Essays on Misallocation and Firm Regulations

Sakai Ando

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018
ABSTRACT

Essays on Misallocation and Firm Regulations

Sakai Ando

This dissertation is a collection of three essays on misallocation and firm regulations. The first chapter investigates how size-dependent firm regulation policies can mitigate misallocation. The second chapter uses the same framework as the first to explore the intuition of a theoretically more subtle concept of misallocation. The third chapter analyzes a more specific firm regulation that targets at financial dealers.

In chapter 1, I study the welfare implications of size-dependent firm regulation policies (SDPs) in the presence of entrepreneurial risks. Although SDP has been considered a source of misallocation, I show that, once entrepreneurial risks are taken into account, SDP might improve efficiency. Quantitatively, I show that, based on French data, removing the SDP leads to output and welfare loss by 1.5% and 1.3%, respectively, in opposition to the output gain reported by the previous literature that abstracts from risks. Qualitatively, I solve an optimal non-linear SDP problem and show that the observed SDP shares certain features with the optimal SDP. The analysis uncovers a novel trade-off between the inefficiencies of the intensive and extensive margins. In extension, it is shown that (1) whether SDPs improve efficiency depends on the level of financial development and (2) capital accumulation and consumption-smoothing motive further justify SDPs.

In chapter 2, which is a joint work with Misaki Matsumura, we use the same competitive entrepreneurship model to investigate the economic intuition of constrained inefficiency caused by uninsurable risks. Although the constrained efficiency of various models has been
studied in the literature, the economic intuition of why the constrained planner’s intervention yields an improvement is usually not available. The competitive entrepreneurship model is particularly suitable for seeing the logic of constrained inefficiency since the structure of the market equilibrium is characterized by the indifference condition instead of the marginal condition. To illustrate this point, we contrast the competitive entrepreneurship model with simple versions of the Aiyagari model and the Krebs model.

In chapter 3, which is also a joint work with Misaki Matsumura, we build a general equilibrium model to analyze the impact of the Volcker rule, a dealer regulation imposed after the financial crisis, on price quality (informativeness and volatility) and its implications on the welfare of market participants. We argue that although price informativeness, volatility and the dealer’s profitability all deteriorate, against conventional wisdom, other market participants are better off due to the dealer’s risk-shifting motive. A static model is used to clarify the main intuition, and the robustness of the welfare results as well as the fragility of the conventional wisdom about price quality are discussed by incorporating dynamics and endogenizing information acquisition.
Contents

List of Figures v

List of Tables vi

1 Size-Dependent Policies and Efficient Firm Creation 1

1.1 Introduction ................................................. 2

1.1.1 Literature ............................................. 9

1.2 Model ....................................................... 11

1.2.1 Risks and efficiency of the laissez-faire economy ........... 16

1.3 Mechanism .................................................. 22

1.3.1 Discussion about the mechanism .......................... 26

1.4 Quantification ............................................... 29

1.4.1 Parameter values ........................................ 31

1.4.2 Results .................................................... 33

1.5 Optimal SDP ................................................ 36

1.6 Financial frictions .......................................... 41

1.6.1 Financial frictions ....................................... 42
# Capital accumulation

- **1.1 Capital accumulation** .......................................................... 48
  - **1.1.1 Equilibrium** ................................................................. 48
  - **1.1.2 Efficiency analysis** ....................................................... 52
- **1.2 Conclusion** ............................................................................ 56

## Constrained Efficiency of Competitive Entrepreneurship

- **2.1 Introduction** ........................................................................... 59
- **2.2 Two simple examples of constrained inefficiency** .................. 61
  - **2.2.1 Aiyagari model** .............................................................. 61
  - **2.2.2 Krebs model** ................................................................. 66
- **2.3 Competitive entrepreneurship model** ...................................... 70
  - **2.3.1 Why is market equilibrium constrained inefficient \( \phi^m \neq \phi^{cp} \)? .......................... 73
  - **2.3.2 Why are entrepreneurs insufficient \( \phi^m < \phi^{cp} \)? ................. 77
- **2.4 Concluding remarks** ............................................................ 79

## Intensive Margin of the Volcker Rule: Price Quality and Welfare

- **3.1 Introduction** ........................................................................... 81
- **3.2 Baseline model** ...................................................................... 87
  - **3.2.1 Environment and definition** ............................................. 87
  - **3.2.2 Characterization of equilibrium** ....................................... 90
  - **3.2.3 Mapping the dealer regulation to the model** .................... 92
  - **3.2.4 Results** ........................................................................... 95
  - **3.2.5 Discussion** ...................................................................... 98
- **3.3 Dynamic model** ................................................................. 99
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1</td>
<td>Environment and definition</td>
<td>100</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Characterization of equilibrium</td>
<td>102</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Policy analysis</td>
<td>103</td>
</tr>
<tr>
<td>3.4</td>
<td>Endogenous information acquisition</td>
<td>106</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Definition of equilibrium</td>
<td>106</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Results and discussion</td>
<td>108</td>
</tr>
<tr>
<td>3.5</td>
<td>Final Remarks</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>111</td>
</tr>
<tr>
<td>A</td>
<td>Appendix to Chapter 1</td>
<td>127</td>
</tr>
<tr>
<td>A.1</td>
<td>Proofs of propositions</td>
<td>127</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Proof of proposition 1.2</td>
<td>128</td>
</tr>
<tr>
<td>A.1.2</td>
<td>Proof of proposition 1.3</td>
<td>130</td>
</tr>
<tr>
<td>A.1.3</td>
<td>Proof of proposition 1.4</td>
<td>133</td>
</tr>
<tr>
<td>A.1.4</td>
<td>Proof of proposition 1.5</td>
<td>135</td>
</tr>
<tr>
<td>A.1.5</td>
<td>Proof of proposition 1.6</td>
<td>139</td>
</tr>
<tr>
<td>A.1.6</td>
<td>Proof of proposition 1.7</td>
<td>140</td>
</tr>
<tr>
<td>A.2</td>
<td>Derivation of Eq. (1.12) and (1.13)</td>
<td>143</td>
</tr>
<tr>
<td>A.3</td>
<td>Non-pecuniary SDP</td>
<td>145</td>
</tr>
<tr>
<td>A.4</td>
<td>MLE estimation</td>
<td>146</td>
</tr>
<tr>
<td>A.5</td>
<td>Intermediate signal</td>
<td>150</td>
</tr>
<tr>
<td>A.6</td>
<td>Constrained efficiency</td>
<td>152</td>
</tr>
<tr>
<td>A.6.1</td>
<td>Constrained efficiency with intermediate signal</td>
<td>152</td>
</tr>
</tbody>
</table>
A.6.2 Constrained efficiency with financial frictions .................................. 155
A.6.3 Constrained efficiency with capital accumulation ........................... 157

B Appendix to Chapter 2 ................................................................. 161

B.1 Proofs of propositions ............................................................... 161
  B.1.1 Proof of proposition 2.3 ....................................................... 161
    B.1.1.1 Cavity of $U(\phi)$ ......................................................... 161
    B.1.1.2 Proof of $\phi^m < \phi^p$ ............................................... 163
  B.1.2 Proof of proposition 2.4 ..................................................... 165

B.2 Heuristic derivation of Eq. (2.18) .............................................. 166

C Appendix to Chapter 3 ................................................................. 168

C.1 Proof of theorem 3.1 ............................................................... 168
C.2 Proof of proposition 3.1 ........................................................... 175
C.3 Proof of proposition 3.2 ........................................................... 176
C.4 Proof of theorem 3.2 ............................................................... 177
C.5 Proof of proposition 3.3 ........................................................... 180
List of Figures

1.1 The number of firms by employment size in France. 3
1.2 Negative correlation between fear of failure rate and entrepreneurship. 4
1.3 Equilibrium system on \((w, \phi)\) plane. 15
1.4 Indifference condition and market clearing condition. 23
1.5 Employment schedule and profit for the laissez-faire economy and the economy with the SDP. 25
1.6 Aggregate output \(Y(\phi)\). 34
1.7 Marginal product as the measure of distortion and the optimal SDP as the percentage of the before-tax profit. 39
1.8 Employment and profit schedule under the estimated threshold SDP, optimal SDP and laissez-faire economy. 41
1.9 Aggregate output \(Y(\phi)\) when market generates too many firms. 42
1.10 Positive correlation of entrepreneurship and financial frictions. 46
1.11 Negative correlation between entrepreneurship and the availability of financing. 47
2.1 The planner’s objective function and the distributions of the entrepreneurs’ profit \(\pi(z, w)\). 77
3.1 Four figures describing the impact of the increase in on equilibrium objects of interest. ................................................................. 96

3.2 Results in a dynamic environment. ............................................ 105

3.3 Results with endogenous information acquisition (EIA). .............. 108

A.1 Counterfactual analysis with intermediate signal. .................... 152
## List of Tables

1.1 Parameter values for counterfactual analysis. ........................................... 31
1.2 Results of the counterfactual analysis. ...................................................... 33
A.1 Comparison ......................................................................................... 149
Acknowledgments

I am deeply indebted to my advisers, Prof. Jennifer La’O and Prof. Jose Scheinkman, who have shaped my taste for research and raised my standard for the clarity of arguments. Aiming to entertain them intellectually has been one of the most important motivations during the Ph.D. program. I am also extremely grateful to Prof. Michael Woodford, whose sharp comments have facilitated my deeper thinking. Other important inputs of my research have been generously provided by Prof. Jushan Bai, Prof. Stephanie Schmitt-Grohe, Prof. Jon Steinsson, Prof. Martin Uribe, Prof. Emi Nakamura, Prof. Paolo Siconolfi, Prof. Pierre Yared, Prof. Takatoshi Ito, and other participants in various seminars.

My student life has been enriched by all my classmates. In particular, discussions with Jean-Jacques Forneron, Yang Jiao, and Xinye Wu were intellectually enjoyable. Among all, I spent most of the time for intellectual production and private entertainment with Misaki Matsumura, who is both my dear wife and my patient coauthor. With special emphasis on my parents who have devotedly supported my Ph.D. life, I thank all people around me for helping my intellectual exploration over the past six years.
Chapter 1

Size-Dependent Policies and Efficient Firm Creation

Sakai Ando
1.1 Introduction

Size-dependent policies (SDPs) that preferentially treat small firms are ubiquitous. For instance, in France, firms that hire more than 50 workers have to pay additional regulatory costs, such as higher tax rates and more stringent labor regulations. Such SDPs naturally create the bunching behavior in the firm size distribution as in Fig.1.1, since some big firms rationally remain small to save regulatory costs. The literature has focused on the misallocation from the bunching behavior, reporting that the removal of SDPs leads to output gain by $0.02 \sim 4\%$ (Gourio and Roys [2014], Garicano et al. [2016]).

In this paper, I argue that our understanding about SDPs can change drastically if we take into account the fact that firm creation is risky. Specifically, I show that, once the analysis incorporates uninsurable entrepreneurial risks, SDPs can generate higher output and welfare, contributing to net efficiency gains despite the remarkable bunching behavior. The key observation is that, when entrepreneurial risks cause insufficient firm creation relative to the Pareto efficient level, SDPs could mitigate the market failure by supporting firm creation. Put differently, if the economy has inefficiencies in the extensive margin, it might be efficiency-enhancing to distort the intensive margin. This is the novel trade-off that this paper uncovers and is at the heart of the following quantitative and qualitative analyses that generate the opposite policy implications to the previous literature.

The key assumption behind the above argument is that entrepreneurial risks deter firm creation. There are several empirical studies that support the assumption. For instance, Hombert et al. [2014] study the French unemployment insurance reform in 2002 and find a positive effect of insuring the downside risks on entrepreneurship. Gottlieb et al. [2017] argue
Figure 1.1: The number of firms by employment size in France. The rise and drop at the employment size of 50 reflect the endogenous reaction of firms to the regulation threshold 50, above which firms are subject to higher tax and heavier labor regulations. The sample includes all the firms in Amadeus 2006 with employment size between 31 and 69.

that the Canadian maternity leave policy reform mitigates the downside risks of losing safe job options and increases entrepreneurship. We can also observe the cross-sectional negative correlation between the downside risks and firm creation as in Fig.1.2.

To illustrate the mechanism and quantify the welfare consequence of SDPs, I extend the standard occupation choice model a la Lucas [1978] (and its descendant Garicano et al. [2016]) by adding entrepreneurial risks. Specifically, in occupation choice, agents choose to become either entrepreneurs or workers. Due to the decreasing returns to scale production technology of entrepreneurs, there is a well-defined number of firms in equilibrium.

The key friction is the uninsurable entrepreneurial risks that agents face as of occupation choice, modeled as the uncertainty about future productivity of the firms that entrepreneurs run. The risks on entrepreneurial productivity generate downside risks since agents are not
Figure 1.2: Negative correlation between fear of failure rate and entrepreneurship. Fear of failure rate is the percentage of 18 – 64 population perceiving good opportunities to start a business who indicate that fear of failure would prevent them from setting up a business. Entrepreneurship is the percentage of the population of 18 – 65 years old who is either a nascent entrepreneur or owner-manager of a new business. The plot pools yearly data 2001 – 2016 for all available countries from 28 to 76 depending on years. The definition of high-income countries and others follow the categorization of the World Bank. Data: Global Entrepreneurship Monitor.

sure whether the business profit is higher or lower than the foregone wage, discouraging risk-averse agents from creating firms. Theoretically, the markets are incomplete due to uninsurable idiosyncratic entrepreneurial risks and therefore the firm creation in the laissez-faire economy is insufficient compared to the Pareto efficient level. This market failure is absent from previous literature and changes the policy implications of SDPs.

The mechanism by which SDPs improve efficiency relies on SDPs facilitating firm creation. Specifically, under the SDPs, big firms pay regulatory costs, so that they limit the number of employees and their profitability deteriorates. The lower labor demand from big firms implies more firm creation in equilibrium since agents are either entrepreneurs or wor-
kers. In other words, SDPs prevent big firms from absorbing human resources that could otherwise be devoted to entrepreneurship. Moreover, the fact that big firms are less profitable makes entrepreneurship less attractive. To offset the resulting increase in the supply of salary workers and reflect weaker labor demand from firms, wage declines in general equilibrium. This is good news for small firms because they do not face regulations and can enjoy lower labor costs. As a consequence, the profitability of small firms improves. From the ex-ante point of view, this means that entrepreneurs face lower downside risks since, even when businesses end up being small or unprofitable, they can earn more than in the laissez-faire economy. As a result of these channels, the economy with SDPs features more firm creation.

The increase in the number of firms can raise output and welfare in the presence of entrepreneurial risks. Intuitively, in the laissez-faire economy with entrepreneurial risks, entrepreneurs produce and consume more on average for the compensation of risk takings. Therefore, SDPs that change the occupation of agents from workers to entrepreneurs can generate positive net output gain, as long as the output loss from bunching is not too large. Accordingly, the higher output and the lower entrepreneurial risks can Pareto improve the laissez-faire economy. Whether the efficiency gain from these channels outweighs the efficiency loss of bunching is a non-trivial question and therefore necessitates the quantification exercise.

To quantify the welfare impact, I conduct the same counterfactual analysis as Garicano et al. [2016] except that my model has entrepreneurial risks. Importantly, I use the same specification and data so that I can isolate the pure implication of entrepreneurial risks. My counterfactual analysis based on French data reveals that removing SDPs decreases output
by 1.5% and welfare in wage unit by 1.27%. The result is in sharp contrast to Garicano et al. [2016] who report that removing SDPs increases output by .02 ∼ 4%.

In terms of implications, the result not only uncovers the channel through which entrepreneurial risks can justify SDPs but also poses a caveat to the misallocation measurement literature in general including Hsieh and Klenow [2009]. In particular, misallocation measurement exercises typically assume that marginal products equalize in the efficient benchmark and measure the deviation from it. The result implies that using such benchmark can be misleading since it might be better to have some wedges among marginal products when the economy has uninsurable entrepreneurial risks. In other words, the trade-off between the inefficiencies of the intensive and extensive margins uncovered in the analysis highlights a limitation of the core logic of the typical misallocation measurement that focuses only on the inefficiency of the intensive margin.

To explore the novel trade-off, I study the non-linear SDP that optimally creates wedges among firms. I solve the optimal SDP problem using the mechanism design approach and show that the optimal SDP and the observed threshold type SDP share qualitative features. Specifically, both the optimal and threshold type SDPs (1) distort medium-sized firms more than the smallest and biggest and (2) feature a larger number of firms compared to the laissez-faire economy. In addition, calibrated at the same parameter values as the quantification exercise, the optimal SDP subsidizes small firms and taxes big firms. In this sense, the observed SDP has efficiency-enhancing properties.

As a novel policy implication, these analyses imply that SDPs can improve efficiency against the current understanding of the literature.

After making the point, I extend my analysis by examining its generality. A natural
question is whether entrepreneurial risks always justify SDPs. I argue that the answer is negative, especially when the laissez-faire economy creates an excessive number of firms. As an example of excessive firm creation, I study the financial frictions that prevent firms from expanding to their profit-maximizing sizes.

It is shown that removing SDPs improves efficiency even when there are entrepreneurial risks if existing firms face severe financial frictions. To be concrete, one can imagine a developing country in which firms cannot expand their employment due to financial frictions. Since firms cannot hire people, those unemployed have to do businesses on their own. As a result, one can observe many small businesses in the economy including people engaged in food truck businesses on the streets. I model such intuition and show that severe financial frictions can lead to too many firms compared to the Pareto efficient level. In this situation, SDPs should be removed since the laissez-faire economy already generates an excessive number of firms in the first place. In fact, the efficiency gains are doubled under excessive firm creation since removing SDPs fixes both the intensive and extensive margins.

The analysis of financial frictions both deepens the policy implication and enrich our understanding about the relationship between financial frictions and entrepreneurship. In terms of the policy implication, the analysis implies that whether SDPs improve efficiency depends on the level of financial development. In financially developed countries with insufficient firm creation, SDPs should be kept, while in countries with immature financial infrastructure, SDPs should be removed.

In terms of financial frictions and entrepreneurship, the claim that entrepreneurial risks cause insufficient firm creation while financial frictions cause the opposite might be surprising since they are both often cited as obstacles to entrepreneurship. The key to understanding the
apparent paradox is to recognize who faces what kind of financial frictions. If entrepreneurs are financially constrained against the fixed cost of entry, the firm creation can be insufficient. However, if existing firms face financial frictions against expansion and the entry cost is a choice variable so that entrepreneurs can start small, the firm creation can be excessive.

Finally, I investigate the implications of dynamic trade-offs. To isolate the implication of inter-temporal decisions in a tractable dynamic environment, I extend the static model by introducing Krep-Porteus-Epstein-Zin preference (Kreps and Porteus [1978], Epstein and Zin [1989]) and capital accumulation a la Krebs [2003]. These modeling techniques allow me to derive a closed form solution despite the heterogeneous-agent incomplete market dynamic model.

It is shown that consumption-smoothing motive provides a further justification for keeping SDPs. In particular, I show that the laissez-faire economy creates insufficient firms, even if agents are risk neutral. The idea is that, when agents prefer smooth-consumption, they want to avoid asset fluctuation, so they do not want to take entrepreneurial risks. Such dynamic trade-off discourages firm creation and therefore justifies SDPs. This result not only confirms the robustness of the analysis in the static model but also highlights a force specific to the dynamic environment.

An implication of the dynamic extension is that SDPs should be kept in an economy with patient agents. This is because patient agents value future consumption more and therefore have stronger consumption-smoothing motives. Theoretically, this reflects the observation that consumption-smoothing motive is controlled not just by the inter-temporal elasticity of substitution but also by the discount factor. Thus, other things being equal, in countries with patient agents such as ones with the culture of patience or long life-expectancy, it is
more likely that removing SDPs exacerbates the market failure of firm creation.

### 1.1.1 Literature

This paper contributes to several strands of literature. First, it provides a novel insight into the conventional wisdom about SDPs. Restuccia and Rogerson [2008], Guner et al. [2008] and Garicano et al. [2016] measure the output gain from removing SDPs. While the details of the models are different across those papers, they share the same feature that the laissez-faire economy without SDPs is efficient. Hence, SDPs decrease output by construction. My contribution is to show that the introduction of entrepreneurial risks might justify SDPs, altering the understanding of SDPs being bad. Another important paper in this strand of literature is Gourio and Roys [2014], which conduct a counterfactual analysis using a firm dynamics model with risky TFP and entry cost. Although they obtain an output loss from removing SDPs, the aggregate consumption and welfare increase, so their policy implication is still to remove SDPs. In contrast, I show a case in which SDPs should be kept.

More broadly, my paper poses a caveat to the misallocation measurement exercises based on marginal product equalization (Hsieh and Klenow [2009], Hsieh et al. [2013]). In particular, I show that, with uninsurable entrepreneurial risks, there is a trade-off between the efficiencies of the intensive and extensive margins, so the efficient benchmark does not necessarily feature marginal product equalization.

The second strand of literature that I contribute studies the efficiency of firm formation. Kihlstrom and Laffont [1979] and Kanbur [1981] are early references for risky firm formation. These papers model both risky occupation choice and non-contingent labor choice so
the resulting market failure is not specific to the friction due to risks. To the best of my knowledge, my paper is the first to study the market failure specific to risky occupation choice. For other environments, Mankiw and Whinston [1986] and Suzumura and Kiyono [1987] are early references that study the optimality of the firm entry in strategic settings. Jaef [2012] studies the optimality of entry in the firm dynamics model of Hopenhayn [1992]. All of them emphasize excessive entry, so the results are opposite to mine.

The third strand of literature that this paper contributes to is the one on financial constraints and entrepreneurship as surveyed by Quadrini [2008] and Buera et al. [2015]. Evans and Jovanovic [1989] is a seminal paper that models liquidity constraint and entrepreneurship. A non-comprehensive list of recent papers includes Buera and Shin [2013], Bohacek and Zubricky [2012], Buera et al. [2011], Cagetti and De Nardi [2009], Kitao [2008] and Meh [2008]. The nature of my exercise is different from these papers. While they are mainly interested in the impact of financial frictions on market equilibrium, I am interested in the impact of financial frictions on the difference between market equilibrium and efficient allocation.

Finally, my paper contributes to dynamic macroeconomics modeling by offering a tractable dynamic general equilibrium model with heterogeneous-agent, incomplete markets and risky occupation choice. The trick behind the tractability is borrowed from Krebs [2003]. Toda [2015] and Gottardi et al. [2016] study the efficiency of the descendants of Krebs [2003]. I extend the framework to include risky occupation choice.
1.2 Model

In this section, I present the baseline model. After discussing the existence and uniqueness of the equilibrium, I use an efficiency analysis to illustrate the market failure that the SDP mitigates. In particular, I show that the laissez-faire economy without SDP generates an insufficient number of firms if and only if there are entrepreneurial risks.

There is a continuum of ex-ante identical risk-averse agents indexed by $i \in [0, 1]$. Each agent is endowed with one unit of indivisible labor that can be spent in running a firm or working for a firm.

If agent $i$ chooses to be a worker, she receives wage $w$ independent of her entrepreneurial productivity $z_i$. If she chooses to be an entrepreneur, she observes her entrepreneurial productivity $z_i$ and then decides the size of the firm measured by the number of employees $n$ to maximize the profit

$$\left[ \pi(z_i, w), n(z_i, w) \right] = \max_{n \geq 0} \begin{cases} 
z_i f(n) - w n & n \leq N \\
z_i f(n) - w \tau n - F & n > N 
\end{cases}$$

(1.1)

where $f$ is the production function, $n(z_i, w)$ is the associated policy function and $(\tau, F, N) \in [1, \infty) \times \mathbb{R} \times \mathbb{R}_+$ is the SDP modeled as the variable and fixed tax that entrepreneurs have to pay when they hire more than $N$. Note that once the firm hires more than $N$ workers, the variable cost $\tau \geq 1$ applies to not just the net additional workers $n - N$ but also all the workers. Therefore, firms have the incentive to shrink the size as in Fig.1.1 even if the fixed component is negative $F < 0$ as long as the total tax payment is positive $(\tau - 1) wN + F > 0$.

Given the payoffs of the two occupations, each agent $i$ observes the signal $s_i$ about the
entrepreneurial productivity $z_i$ and chooses the occupation that maximizes her expected utility. The joint distribution of the signal and productivity $(s, z)$ is exogenously given and denoted by $G$. Formally, each agent $i$ solves

$$
e (s_i, w) = \arg \max_{e \in [0,1]} e \mathbb{E}\left[u(\pi(z_i, w)) | s_i\right] + (1 - e) u(w)$$

(1.2)

where $u$ is the utility function, and $\mathbb{E}[\cdot | s_i]$ is the expectation with respect to the productivity $z_i \in \mathbb{R}^+$ conditional on the observed signal $s_i \in \mathbb{R}$. If the agent chooses $e = 1$, she becomes an entrepreneur, while $e = 0$ indicates that she becomes a worker.

I make two observations about the occupation choice problem. First, note that the choice variable $e$, representing whether to become an entrepreneur, can take continuous values $e \in [0,1]$. This formulation allows agents to take mixed strategy when they are indifferent between the two occupations. Second, the occupation choice is risky because the agent has to decide the occupation before observing the productivity $z_i$. Since the salary job is risk-free, depending on the realization of the productivity $z_i$, entrepreneurs might consume less than workers in the equilibrium. In this sense, entrepreneurship involves downside risks.

Finally, the wage $w$ clears the labor market

$$1 - \int e (s_i, w) \, di = \int e (s_i, w) n (z_i, w) \, di.$$  

(1.3)

where the left hand side is the aggregate labor supply and the right hand side denotes the aggregate labor demand from firms.

The following definition summarizes the description of the equilibrium as well as other
objects of interests.

**Definition 1.1.** Fix the fundamentals \((u, f, G)\) and the SDP \((\tau, F, N)\). The set of wage, occupation choice and production decisions \(\{w, e, \pi, n\}\) is an equilibrium if it satisfies (1.1), (1.2) and (1.3). The number of firms and the aggregate output in the equilibrium is defined as

\[
\phi = \int e(s_i, w) \, di, \quad Y = \int e(s_i, w) z_i f(n(z_i, w)) \, di. \tag{1.4}
\]

The laissez-faire economy is defined as the one with \((\tau, F) = (1, 0)\).

Throughout the paper, I attach LF and SDP to the equilibrium objects whenever the distinction of the laissez-faire economy and the economy with the SDP needs to be made explicit.

To ensure that the equilibrium is well-behaved, I make three assumptions on the fundamentals \((u, f, G)\). First, utility and production functions \((u, f)\) are strictly increasing, strictly concave and satisfy the Inada condition. As will be clear, the strict concavity of the production function ensures the optimal level of firm creation. Second, the joint distribution of the signal productivity \(G(s, z)\) is continuous, has bounded productivity, i.e. \(P(0 < z_{\text{min}} \leq z \leq z_{\text{max}} < \infty) = 1\) for some \((z_{\text{min}}, z_{\text{max}})\), and reflects positively informative signals, i.e., the conditional distribution \(G(z|s)\) first-order stochastically dominates \(G(z|s')\) whenever \(s > s'\). The continuity allows the system to adjust continuously, the bounded productivity saves the complexity of infinite utility, and the positively informative signal provides a normalization so that the higher signal one observes, the more likely one gets higher productivity. Finally, I assume the signal and productivity \((s_i, z_i)\) are drawn \(i.i.d.\) from \(G\). The \(i.i.d.\) assumption makes it possible to invoke the law of large numbers.
I call these assumptions \((A)\). Under these assumptions, the existence and the uniqueness can be guaranteed as stated in the next proposition. I also show the proof since it clarifies the structure of the equilibrium.

**Proposition 1.1.** Suppose that the fundamentals \((u, f, G)\) satisfy the assumptions \((A)\). An equilibrium exists and is unique almost surely.

**Proof.** Given the signal structure, the occupation choice follows a threshold strategy \(e(s_i) = 1_{s_i \geq \bar{s}}\). The "almost surely" qualification reflects the indeterminacy of optimal occupation for the marginal entrepreneur \(s_i = \bar{s}\). Given the individual occupation choice \(e(\cdot)\), the market equilibrium can be characterized by the equilibrium wage \(w\) and the threshold \(\bar{s}\) that satisfy the following indifference condition and the market clearing condition

\[
\mathbb{E}[u(\pi(z, w)|s = \bar{s})] = u(w), \quad (1.5)
\]

\[
\mathbb{E}[n(z, w)1_{s \geq \bar{s}}] = G_s(\bar{s}). \quad (1.6)
\]

where Eq. (1.3) invokes the law of large numbers a la Uhlig [1996].

To show the existence and the uniqueness, note that, for a fixed SDP \((\tau, F, N)\), the indifference condition (1.5) specifies a positive relationship between the wage \(w\) and the threshold \(\bar{s}\), while the market clearing condition (1.6) provides a negative relationship. Both of these relationships are continuous since \(n(z, w)\) jumps at most at one point. Moreover, according to the market clearing condition, \(w \to \infty\) as \(G(\bar{s}) \to 0\), and \(w \to 0\) as \(G(\bar{s}) \to 1\). Therefore, the two loci must cross each other once.
relationships between the wage $w$ and the number of firms $\phi = 1 - G_s(s)$. One can graphically see them in Fig. 1.3. First, the indifference condition implies that if the wage increases, more people find it attractive to become workers, so the number of firms declines. This is the labor supply side intuition from the individual perspective and generates the negative relationship between the wage and the number of firms. Second, the market clearing condition implies that, if the wage increases, firms’ labor demand declines, so there will be fewer workers, ending up with more firms. This labor demand side story gives the positive relationship between the wage and the number of firms. Since these two forces bring the system to the opposite directions, there exists a unique equilibrium irrespective of the specific form of the productivity and the signal structure $G$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{IC_and_MC_on_w_phi_plane.png}
\caption{Equilibrium system on $(w, \phi)$ plane. The indifference condition gives a negative relationship, while the market clearing condition generates a positive relationship.}
\end{figure}
1.2.1 Risks and efficiency of the laissez-faire economy

The key to understanding the role of the SDP is the market failure due to entrepreneurial risks. This section defines the riskiness of entrepreneurship and shows that the laissez-faire economy without the SDP is inefficient if and only if there are entrepreneurial risks. In particular, I highlight the insufficient firm creation in the presence of entrepreneurial risks.

The riskiness of entrepreneurship is defined by how informative the signal $s$ is about the productivity $z$. If the signal $s$ is very informative about entrepreneurial productivity $z$, agents can accurately forecast future profits when they become entrepreneurs. In contrast, if agents have uninformative signal $s$, becoming entrepreneurs may involve substantial downside risks.

One way to define informativeness mathematically is to use the information sets. If the information sets of the two random variables are identical, i.e. the sigma algebras are identical $\sigma s = \sigma z$, there are no entrepreneurial risks. In contrast, if the signal is independent of the productivity $\sigma s \perp \perp \sigma z$, there are full risks. Although this definition has generality, it does not come with a natural framework to think about the intermediate cases. Thus, the current section uses a weaker condition to define riskiness.

**Definition 1.2.** The riskiness of entrepreneurship for those who observe $s$ is defined as the conditional variance $V(z|s) \in [0, V(z)]$. $V(z|s) = 0$ corresponds to no risks, while $V(z|s) = V(z) > 0$ corresponds to full risks.

One can use weaker conditions, defining no risks as the signal structure where each agent knows the best occupation for sure, i.e., $P(\pi(z,w) \geq w|s) \in \{1, 0\}$ for all $s$, and full risks as the signal structure where the signal is not informative about the occupation choice, i.e. $P(\pi(z,w) \geq w|s) = P(\pi(z,w) \geq w)$ for all $s$. This definition allows intuitive intermediate
cases for some signal structure and is adopted in section 1.4.

In any case, the economy with no risks is characterized by \((\bar{z}, w)\) that satisfies the following indifference and market clearing conditions

\[
\pi(\bar{z}, w) = w, \quad \mathbb{E}[n(z, w) 1_{z \geq \bar{z}}] = G_z(\bar{z}),
\]

and the economy with full risks is characterized by \((w, \phi)\) that satisfies another set of indifference and market clearing conditions

\[
\mathbb{E}u(\pi(z, w)) = u(w), \quad \phi\mathbb{E}n(z, w) = 1 - \phi.
\]

One can see that, without risks, the utility function disappears from the equilibrium system. With full risks, the equilibrium wage is determined by the indifference condition, and the number of firms is determined by the market clearing condition. Graphically, the full-risk case requires the indifference condition in Fig.1.3 to be horizontal for the interior number of firms. In terms of the literature, the two cases correspond to Lucas [1978] and Kanbur [1979a], and Garicano et al. [2016] use the former to estimate the efficiency loss due to the SDP.

Given the riskiness, I can discuss the efficiency of the laissez-faire economy. Once the market failure of the laissez-faire economy is understood, I describe the mechanism by which the SDP improves efficiency in section 1.3.

To discuss the efficiency, I define the planner’s problem who can choose allocations at each state but faces the same informational constraints as individual agents. Formally,
a state of the economy is a collection of the productivity realizations for all agents \( \omega = \{ z_i \}_i \in \Omega = [z_{min}, z_{max}]^{[0,1]} \). The planner chooses the contingency plan of the consumption \( c = \{ c_i(\omega) \}_{i \in [0,1], \omega \in \Omega} \), the employment schedule when agents become entrepreneurs \( \{ n_i(\omega) \}_{i \in [0,1], \omega \in \Omega} \) and the allocation of occupation \( e = \{ e(s_i) \}_i \). Note that the consumption and employment are contingent on each state, but the occupation has to be measurable with respect to the signal \( s \) since that is the informational constraint agents face. Such informational constraint is standard in the literature of efficiency analysis with informational frictions. (Angeletos and Pavan [2007])

Given a measurable Pareto weight \( i \mapsto \Lambda_i \), the planner maximizes the weighted sum of the expected utility

\[
U(c, \Lambda) = \int \mathbb{E}_\Omega u(c_i(\omega)) \, d\Lambda_i
\]  

where \( \mathbb{E}_\Omega \) denotes the expectation over the state \( \Omega \) constructed from \( G_z \).  

Since the economy consists of heterogeneous agents differentiated by the observed signal \( \{ s_i \}_i \), I do not take a stand on the Pareto weight such as utilitarian or Rawlsian, and instead, focus on the production side. Note that since the economy does not have aggregate uncertainty and production in one firm is independent of the productivity of other firms, I can restrict the employment schedule to be measurable with respect to productivity, \( n = \{ n(z) \}_z \). Furthermore, since the planner maximizes output, the allocation of occupation is a threshold rule \( e(s_i) = 1_{s_i \geq \bar{s}} \) without loss of generality. This is equivalent to choosing the number of firms \( \phi = 1 - G_s(\bar{s}) \). Hence, the planner’s problem can be defined as follows.

**Definition 1.3.** Fix the fundamentals \( (u, f, G) \). A set of consumption, employment schedule,
output and number of firms \( \{c,n,Y,\phi\} \) is Pareto efficient if it solves
\[
\max_{c,n,Y,\phi} U(c, \Lambda) \quad \text{s.t.} \quad \begin{cases}
\int c_i(\omega) \, di = Y \\
Y = \phi \mathbb{E} \left[ zf(n(z)) \, \vert \, s \geq G_s^{-1}(1 - \phi) \right] \forall \omega \in \Omega \\
\phi + \phi \mathbb{E} \left[ n(z) \, \vert \, s \geq G_s^{-1}(1 - \phi) \right] = 1
\end{cases}
\] (1.10)

for some Pareto weight \( \Lambda \).

The first constraint requires that individual consumption adds up to the aggregate output. Since the economy does not have aggregate uncertainty, it implies that the planner can provide full insurance, \( c_i(\omega) = c(\omega') \) for all \( \omega, \omega' \in \Omega \). An immediate implication is that the planner chooses employment schedule \( n \) and the number of firms \( \phi \) to maximize output. The second constraint illustrates the production technology as a function of \( (n, \phi) \). There are \( \phi \) firms and each of them produces \( \mathbb{E} \left[ zf(n(z)) \, \vert \, s \geq G_s^{-1}(1 - \phi) \right] \) on average. The conditional expectation reflects the entrepreneurs selection \( \bar{s} = G_s^{-1}(1 - \phi) \). The third constraint describes the human resource constraint. \( \phi \) agents become entrepreneurs and each of them hire \( \mathbb{E} \left[ n(z) \, \vert \, s \geq G_s^{-1}(1 - \phi) \right] \) workers, which have to add up to the total population 1.

In general, the planner’s solution depends on the Pareto weight. The production side, however, can be uniquely determined independently of the Pareto weight.

**Proposition 1.2.** Suppose that the fundamentals \( (u, f, G) \) satisfy the assumptions \( (A) \). Then, there is a unique interior solution \( \phi^P \in (0,1) \) to the planner’s problem (1.10).

**Proof.** See Appendix A.1.1. \( \square \)

\(^2\)Although all variables are functions of states, the assumption of no aggregate uncertainty saves the notation for production side.
To understand the planner’s trade-off that pins down the interior solution, let \( Y(\phi) \) denote the aggregate output as a function of the number of firms, obtained by maximizing out employment schedule \( \{n(z)\}_z \) from the last two constraints of (1.10). To be concrete, let the production function be Cobb-Douglas \( f(n) = n^\alpha \) with \( \alpha \in (0,1) \). The aggregate output \( Y(\phi) \) takes the following form

\[
Y(\phi) = A(\phi) \phi^{1-\alpha} (1-\phi)\alpha, \quad A(\phi) = \left( E \left[ z^{1-\alpha} | s \geq G_s^{-1}(1-\phi) \right] \right)^{1-\alpha}. \tag{1.11}
\]

One can see the two channels through which the number of entrepreneurs impacts the aggregate output.

The first channel is the average productivity \( A(\phi) \). It describes a typical concern about entrepreneurship promotion. If the number of entrepreneurs increases \( \phi \rightarrow \), since marginal entrepreneurs is less productive \( G_s^{-1}(1-\phi) \), the average productivity declines \( A(\phi) \). However, the average productivity \( A(\phi) \) is not the main technological trade-off that pins down the unique interior solution. After all, the average productivity is bounded \( A(\phi) \in [z_{\min}, z_{\max}] \), and does not even react to the number of firms \( \phi \) when the signal \( s \) is uninformative about the productivity \( z \).

The second channel is the allocation of occupations \( \phi^{1-\alpha} (1-\phi)\alpha \). This is the key trade-off that pins down the unique interior solution. When most agents are workers \( \phi \rightarrow 0 \), the small number of firms employs a large number of workers. Due to the decreasing returns of scale assumption on the production technology \( f \), the worker’s marginal product is low. In this case, dividing the firm into two and shifting some of the workers to higher marginal product activities raises the aggregate output. However, such firm creation is not cost-less
since it requires a decrease in the number of workers engaged in production. In the limit, if everyone becomes an entrepreneur $\phi \to 1$, no workers make production, resulting in 0 aggregate output despite the worker’s marginal product being $\infty$.

This trade-off specified by the second channel is present even if the productivity distribution $G_z$ is degenerate or the signal is not informative $s \perp z$. In these cases, the optimal number of firms is $\phi^P = 1 - \alpha$. It also differentiates the occupation choice models from the firm dynamics models following Hopenhayn [1992], in which labor supply is fixed and no entrepreneurs are needed to create additional firms.

Now we are ready to state the efficiency result. Let $\phi^{LF}$ and $\phi^P$ be the number of firms in the laissez-faire economy and the planner’s solution. The following proposition states that entrepreneurial risks create the market failure of insufficient firm creation. Note that the assumptions on the riskiness $G(z|s)$ affect both the market equilibrium and the planner’s solution.

**Proposition 1.3.** Suppose that the fundamentals $(u, f)$ satisfy the assumptions (A).

1. If there are no entrepreneurial risks, i.e., $V(z|s) = 0$ for all $s$, the laissez-faire economy is Pareto efficient.

2. If there are entrepreneurial risks, i.e., $V(z|s) > 0$ for all $s$, the laissez-faire economy is not Pareto efficient. In particular, the laissez-faire economy creates insufficient number of firms $\phi^{LF} < \phi^P$.

*Proof.* See Appendix A.1.2.

The results are not surprising if one notices that the economy without risks has complete
markets so the analogy of the first welfare theorem holds, while the economy with risks features incomplete markets and uninsurable idiosyncratic risks. As is standard in the literature of incomplete markets, the result can be extended to constrained inefficiency as shown in Appendix A.6.1 and discussed in details in Ando and Matsumura [2017].

The intuition of the results is transparent if one considers a marginal increase of firms and aggregate output. Without risks, the entrepreneur with productivity \( z \) produces \( \pi(z, w) \) and workers produce \( w \). If one worker changes the occupation to entrepreneur, the net output gain is \( \pi(z, w) - w \). In the laissez-faire economy (1.7), the marginal entrepreneurs are indifferent to the workers \( \pi(\bar{z}, w) = w \). Therefore, increasing firms results in zero net output gain \( \pi(\bar{z}, w) - w = 0 \). With risks, the marginal entrepreneurs produce more than workers \( \mathbb{E}[\pi(z, w) | s = \bar{s}] > w \) in the laissez-faire economy. This is because for risk-averse agents to become entrepreneurs, they have to be able to produce and consume more than workers on average. As a result, switching the marginal workers into entrepreneurs raise aggregate output by \( \mathbb{E}[\pi(z, w) | s = \bar{s}] - w > 0 \).

1.3 Mechanism

I have shown in the previous section that Pareto improvement is only possible if there are entrepreneurial risks. This section uses the full-risk model to magnify the role of risks and describe how the SDP could make Pareto improvement. After describing the mechanism, I discuss its plausibility.

The key observation is that the SDP may increase the number of firms by (1) preventing big firms from absorbing potential entrepreneurs and (2) reducing the entrepreneurial risks
Figure 1.4: Indifference condition and market clearing condition.

through general equilibrium. To see the first channel, note that the SDP imposes regulatory costs on big firms, so they hire fewer workers. Since agents are either entrepreneurs or workers, the decline in the number of workers implies the rise in the number of firms increases in equilibrium. This is a partial equilibrium impact with fixed wage and is illustrated as the shift of the market clearing condition in Fig.1.4.

In general equilibrium, wage decreases. To see this, note that the SDP makes entrepreneurship less attractive in partial equilibrium, i.e., $\pi^{SDP}(z, w^{LF}) \leq \pi^{LF}(z, w^{LF})$ for all $z \in [z_{min}, z_{max}]$. Since the inequality is strict for big firms, the indifference condition (1.8) implies that the wage declines $w^{SDP} < w^{LF}$. Intuitively, since entrepreneurship is less attractive, agents want to become workers. As a result of increase in worker supply, the wage decreases. This force is illustrated as the shift of the indifference condition in Fig.1.4.\(^3\)

\(^3\)When the signal is informative, the lower labor demand from big firms also pushes down the wage. Mathematically, the indifference condition is not horizontal, so the shift of the market clearing condition itself causes wage decline.
The wage decline mitigates the downside risks of entrepreneurship. This is because, from the ex-ante point of view, even when businesses end up being small or unprofitable, entrepreneurs do not have to pay the regulatory costs but can enjoy cheaper labor costs. The mitigation of the downside risks is illustrated in the right figure of Fig.(1.5). In partial equilibrium, the profit in the laissez-faire economy is higher for all productivity levels. However, such profit shift makes entrepreneurship less attractive. To regain equilibrium, the wage has to decline. As a result, the small firms that do not face regulation can hire workers cheaply, resulting in higher profit.

Both direct regulations on big firms and indirect subsidies to small firms contribute to firm creation. The former prevents big firms from absorbing potential entrepreneurs, and the latter reduces the downside risks of entrepreneurship. Thus, the SDP may support firm creation.

The increase in the number of firms may lead to output increase. The intuition can be obtained by the perturbation argument. I show the results and relegate the derivation to Appendix A.2. Suppose the government implements a small value of SDP around the laissez-faire $T = (\tau, F) \approx (1, 0)$. The marginal impact on the aggregate output is

$$\partial_T Y = \partial_T \phi (\mathbb{E}_{\pi}(z, w) - w). \quad (1.12)$$

The expression illustrates that the policy impact around the laissez-faire economy is the multiplication of how many firms the policy increases $\partial_T \phi$ and how much additional output each firm produces $\mathbb{E}_{\pi}(z, w) - w$. Intuitively, since entrepreneurs produce $\mathbb{E}_{\pi}(z, w)$ and workers $w$, if an agent changes the occupation from worker to entrepreneur, the net gain is
Figure 1.5: Employment schedule $n(z,w)$ and profit $\pi(z,w)$ for the laissez-faire economy and the economy with the SDP. PE denotes partial equilibrium where the wage is fixed at laissez-faire level $w^{LF}$. GE means general equilibrium where the wage is $w^{SDP}$.

$\mathbb{E}\pi(z,w) - w$ on average. This term is positive since in the laissez-faire economy, risk aversion requires entrepreneurs to produce and consume more $\mathbb{E}\pi(z,w) - w > 0$ in equilibrium for the compensation of the risk-taking. Therefore, the SDP that increases the number of firms $\partial_T \phi > 0$ increases the aggregate output.$^4$

The increase in output and the reduction of risks may lead to Pareto improvement if the government returns the tax revenue. Note that, since agents are ex-ante identical in the full-risk model (1.8), the welfare can be measured by the worker’s consumption. As a result, if the government throws away the tax revenue, all agents are worse off due to the lower wage. However, if the government has enough tax revenue and return it to agents, the disposable

\[ \mathbb{E}\pi(z,w) - w \]

since $\pi(\bar{z},w) = w$ in laissez-faire economy, the net output gain is $\partial_T Y = 0$ even though the SDP increases the number of firms.

$^4$The output increase does not happen in no-risk case. To see this, note that the same perturbation argument leads to

\[ \partial_T Y = -g_z(\bar{z}) \partial_T \bar{z} \cdot (\pi(\bar{z},w) - w). \]
wage can increase. In this sense, Pareto improvement is possible if the government raises enough revenue and return it.

Each of the above steps involves non-trivial assumptions and necessitates the quantification exercise in section 1.4. First, the increase in the number of firms hinges on the wage not declining too much. If wage declines to \( w \approx 0 \), firms can hire many workers, so the number of entrepreneurs decreases. In other words, the general equilibrium impacts on the system through wage cannot be larger than the partial equilibrium impacts. As is shown in section 1.4, this is true for standard parametric specifications. Second, the output increase is discussed around the laissez-faire economy. This is to illustrate the intuition of the efficiency gain. For the SDP far from the laissez-faire, there is efficiency loss due to the bunching. Therefore, whether the output increases or not depends on the balance of the two forces, and has to be determined numerically. This necessitates the quantification exercise in section 1.4. Third, the Pareto improvement requires the government to raise enough revenue. Although the SDP increases output and reduces risks for each entrepreneur, the number of entrepreneurs also increases. Since the entrepreneurs need to consume more on average to obtain the same utility as workers, whether the SDP can increase enough output to Pareto improve all agents is not obvious. Again, the non-triviality necessitates the quantification exercise in section 1.4.

1.3.1 Discussion about the mechanism

The mechanism of welfare improvement relies on debatable assumptions and generates seemingly counter-intuitive implications. In this section, I provide a discussion for each of
First, the logic that workers change occupations to become entrepreneurs might naturally generate concerns about unemployment. Although the model does not address unemployment, I make two heuristic observations. First, there are many countries that have both SDPs and low unemployment rates. Therefore, at least in the long run, SDPs and unemployment can be considered separate issues. Second, the thought experiment that I am interested in is the removal of the SDP. Thus, the relevant trade-off is the lower unemployment rate in the short run and the lower aggregate output in the long run. Therefore, the social cost associated with unemployment is not an issue in both the long and short run in the current context.

Second, the output increase depends on entrepreneurs producing more on average $E\pi(z, w) > w$ than workers in the laissez-faire economy. Since many countries adopt SDPs, it is not easy to empirically test whether $E\pi(z, w) > w$ holds in the unobservable laissez-faire economy. That being said, it is informative to understand the debate about entrepreneurial returns in general, assuming $E\pi^{SDP}(z, w^{SDP}) > w^{SDP}$ implies $E\pi^{LF}(z, w^{LF}) > w^{LF}$. The seminal papers that claim the inequality $E\pi(z, w) > w$ is violated are Hamilton [2000] and Moskowitz and Vissing-Jorgensen [2002]. Their studies have generated a literature that tries to explain why people choose to become entrepreneurs in the first place (e.g. Vereshchagina and Hopenhayn [2009]).

More recent papers argue against $E\pi(z, w) > w$. Levine and Rubinstein [2017] argue that the definition of entrepreneurs used in the literature many not be appropriate. If entrepreneurs are defined to be incorporated self-employed, the mean return is much higher than salaried workers. Manso [2016] points out that the estimations in the literature might
be biased due to the usage of cross-sectional data to compute the mean of entrepreneurial earnings. Manso [2016] reports that, once the option value of experimentation is incorporated, the mean lifetime earnings of entrepreneurs are higher than those of salary workers in the U.S. Given that there might be many other factors that affect the returns on the two occupations, I explore the implications of the standard occupation choice model in this paper.

Third, some might feel the mechanism is counter-intuitive because output increases even though big firms hire fewer workers under the SDP. One way to understand the logic intuitively is to consider an extreme example. Suppose one company hires the entire population. Even if the firm is productive, this is not necessarily a good thing if it has decreasing returns to scale. This is because eventually, the marginal product of employees declines. What the SDP does is reduce the firm’s employees and facilitate their engagement in firm creation. Since the workers in the new firms have higher marginal products, an increase in the number of firms is an improvement in production efficiency. I discuss when this intuition breaks in section 1.6.

Finally, the Pareto improvement is based on tax return. Since many SDPs are non-pecuniary such as working hours regulations, there might not be tax revenue in reality. In this case, the fair comparison is (1) the economy with SDP and (2) the laissez-faire economy with the same amount of non-distortionary regulation costs. In Appendix A.3, I take the extreme by assuming the SDP $(\tau, F)$ is completely non-pecuniary, and show that the economy with the SDP (1) still generates higher welfare than the counterfactual (2).
1.4 Quantification

This section quantifies the welfare impact of removing the SDP. The main result is that removing the SDP leads to lower output and welfare. The result of output decline is the opposite to the previous literature and highlights the potential importance of the channel described in section 1.3.

To isolate the role of entrepreneurial risks, I closely follow the setup of Garicano et al. [2016]. In particular, I impose the parametric assumptions on the fundamentals, with utility and production functions being constant relative risk aversion and Cobb-Douglas

\[
u(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma}, \quad \gamma > 0, \quad f(n) = n^\alpha, \quad \alpha \in (0, 1)
\]  

and the distribution of the productivity being the power law

\[
g_z(z) = \frac{1 - \beta_z}{z_{\max}^{1 - \beta_z} - z_{\min}^{1 - \beta_z}} z^{-\beta_z}, \quad z_{\min} \leq z \leq z_{\max}.
\]  

This productivity distribution leads to a broken power law of the firm size distribution as detailed in Appendix A.4.

I focus on the full risks model, (1.8), and relegate a specific intermediate signal case in Appendix A.5. In this way, the quantitative exercise in this section is the exact opposite polar case of Garicano et al. [2016] with the same set of fundamentals. Therefore, one can transparently see the implications of the entrepreneurial risks without being affected by the specific choice of the signal structure.

Formally, the factual economy with the SDP \((\tau, F, N)\) and the non-distortionary income
tax \( t \) gives the equilibrium \((w^{SDP}, \phi^{SDP})\) that satisfies the indifference condition

\[
\mathbb{E}u \left((1 - t) \pi (z, w)\right) = u \left((1 - t) w\right), \tag{1.16}
\]

profit maximization (1.1), and the market clearing condition in (1.8). The income tax \( t \) is non-distortionary because, when the utility function is CRRA, \((1 - t)\) drops out of the equilibrium system as is clear from (1.16). As a result, all the equilibrium objects except for welfare are independent of the income tax rate \( t \). To determine the welfare, I set the income tax rate \( t = t(\tau, F) \) by imposing the balanced budget constraint

\[
\phi^{SDP} \mathbb{E} \left\{ \left[ (\tau - 1)w^{SDP} n(\tau, w^{SDP}; \tau, F) + F \right] \mathbb{1}_{n(w^{SDP}; \tau, F) > N} \right\} + t \left\{ \phi^{SDP} \mathbb{E}\pi (z, w^{SDP}; \tau, F) + \left(1 - \phi^{SDP}\right) w^{SDP} \right\} = 0. \tag{1.17}
\]

I compare the factual economy with the counterfactual laissez-faire economy where \((t, \tau, F) = (0, 1, 0)\). Specifically, I compare aggregate output, welfare, number of firms, and wage in the two economies. Note that the welfare is given by the disposable income \((1 - t) w^{SDP}\) and \(w^{LF}\).

There are two differences from Garicano et al. [2016]. The first is the existence of entrepreneurial risks, and the second is the welfare criterion. The welfare criterion of Garicano et al. [2016] is aggregate output minus tax revenue. To make the comparison consistent, I also report output minus tax revenue in section 1.4.2. An alternative welfare specification assuming all the SDP being non-pecuniary is available in Appendix A.3. Irrespective of the specification of the welfare criterion, all other equilibrium objects \((Y, w, \phi)\) are directly
<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>2</td>
<td>CRRA parameter</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.8</td>
<td>Cobb-Douglas parameter</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.023</td>
<td>marginal tax</td>
</tr>
<tr>
<td>$F/w$</td>
<td>-.941</td>
<td>normalized fixed tax</td>
</tr>
<tr>
<td>$\beta_z$</td>
<td>5</td>
<td>power law parameter of $G_z$</td>
</tr>
<tr>
<td>$z_{max}$</td>
<td>8.07</td>
<td>productivity upper bound</td>
</tr>
<tr>
<td>$z_{min}$</td>
<td>1.4</td>
<td>productivity lower bound</td>
</tr>
</tbody>
</table>

Table 1.1: Parameter values for counterfactual analysis. All parameters are taken from Garicano et al. [2016] except for $(\gamma, z_{min})$. Wage $w$ is not identified so is normalized to $w = 1$ without loss of generality.

1.4.1 Parameter values

To conduct the counterfactual analysis, I need to specify the value of the parameters. In this section, I show the parameter values and describe the steps. The main takeaway is that I can use the same parameter values as Garicano et al. [2016], and therefore all the differences in the results are driven by the existence of entrepreneurial risks, not other elements such as the data used and specifications of the model.

Table 1.1 summarizes the choice of the parameter values based on the same French data used by Garicano et al. [2016]. Note that the fixed tax $F$ is negative. This does not mean the firms at size $N$ get subsidies since the variable tax $\tau$ applies to all workers, not just the net additional workers $n - N$. Indeed, the tax payment for firms at $N$ is $(\tau - 1)wN + F = .186w > 0$.

The main parameters of interest is the SDP $(\tau, F, N)$. The threshold $N$ is determined
by the law so there is no other choice than \( N = 49 \). The variable and fixed components 
\((\tau, F)\), as pioneered by Garicano et al. [2016], can be estimated jointly with the power law 
parameter \( \beta_z \) by using the data of firm size distribution. To be more concrete, one can derive 
the equilibrium firm size distribution \( n \mapsto G_n(n; \tau, F, \beta_z) \) that follows a broken power law 
as detailed in Appendix A.4, and use the maximum likelihood method to fit it to the firm 
size distribution data \( \{n_i\} \) by choosing \((\tau, F, \beta_z)\) and other auxiliary parameters. As a trick 
to cleanly estimate the parameters, Garicano et al. [2016] truncate the data and likelihood 
function \((n \leq 10 \text{ and } n \geq 10,000)\) so that the two tails do not impact the estimation.

In principle, since I use a different model with entrepreneurial risks, I need to recollect the 
same data and redo the same estimation. However, I can use the same estimates as Garicano 
et al. [2016] because the implied firm size distributions after truncation are identical. To see 
this, recall that the two models, (1.7) and (1.8) share the same fundamentals \((u, f, G_z)\). The 
only implication of the entrepreneurial risks on the firm size distribution \( G_n(n) \) is whether 
low productivity agents become entrepreneurs or not. With no risks, only high productivity 
agents \( \{i : \pi(z_i, w) \geq w\} \) become entrepreneurs, while with full risks, all productivity levels 
are possible. However, the firm size distributions \( G_n(n) \) in the two models are both broken 
power laws. Since the broken power law is invariant to truncation, the firm size distributions 
after truncation are identical, and therefore generate the same estimates.\(^5\)

Other parameters follow the calibration and normalization of Garicano et al. [2016]. For 
instance, the equilibrium wage is unidentified, so is normalized to be 1. The parameter that 
is not in Garicano et al. [2016] is the risk aversion \( \gamma \). I choose the standard number \( \gamma = 2 \) and

\(^5\)The likelihood function in Garicano et al. [2016] also incorporates measurement error. Although this 
complication might potentially create discrepancies between the estimates, I show they are numerically 
negligible in Appendix A.4 using Amadeus data and the models with and without entrepreneurial risks.
### Table 1.2: Results of the counterfactual analysis.

<table>
<thead>
<tr>
<th>Equilibrium objects</th>
<th>Full risks</th>
<th>No risks</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln Y^{LF} - \ln Y^{SDP}$</td>
<td>-1.5%</td>
<td>.02%</td>
<td>output</td>
</tr>
<tr>
<td>$\ln \phi^{LF} - \ln \phi^{SDP}$</td>
<td>-8%</td>
<td>-7.2%</td>
<td>number of firms</td>
</tr>
<tr>
<td>$\ln Y^{LF} - \ln (Y^{SDP} - \text{tax})$</td>
<td>-.3%</td>
<td>1.3%</td>
<td>output - tax</td>
</tr>
<tr>
<td>$\ln w^{LF} - \ln {(1 - t) w^{SDP}}$</td>
<td>-1.25%</td>
<td>NA</td>
<td>welfare</td>
</tr>
<tr>
<td>$\ln w^{LF} - \ln w^{SDP}$</td>
<td>.002%</td>
<td>1.8%</td>
<td>wage</td>
</tr>
</tbody>
</table>

The results are summarized in Table 1.2. The first column shows the results of the counterfactual analysis based on my model with full entrepreneurial risks. The second column is a copy from Garicano et al. [2016], which is based on the model with no entrepreneurial risks.

1.4.2 Results

This section presents the results of the counterfactual analysis. It is shown that the output is 1.5% lower and the welfare in wage-unit is 1.27% lower in the laissez-faire economy without the SDP.

The results are summarized in table 1.2. The first column shows the results of the counterfactual analysis based on my model with full entrepreneurial risks. The second column is a copy from Garicano et al. [2016], which is based on the model with no entrepreneurial risks.

The first row shows that the factual economy has the higher output than the counterfactual economy. This might be surprising if one just sees the marginal products. Specifically, the marginal products equalize among all firms in the laissez-faire economy, yet the laissez-faire economy produces less than the factual economy with wedges across firms. Fig.1.6
describes the intuition behind the result. The red line represents the production possibility frontier of the economy with the SDP $Y^{SDP}(\phi)$ as a function of the number of firms $\phi$. For each fixed number of firms, the removal of the SDP improves production efficiency so the production possibility frontier of the laissez-faire economy is located above. Therefore, for a given number of firms, removing SDP does improve resource allocation. However, removing the SDP exacerbates the market failure of firm creation. In particular, as discussed in section 1.3, big firms suck up human resources from entrepreneurial activities and the entrepreneurship itself is riskier. As a result, the number firms declines as shown in the second row. The counterfactual analysis shows that the efficiency loss of firm creation can be larger than efficiency gain from marginal product equalization.\footnote{This does not happen in Guarino et al. [2016] because the laissez-faire economy is at the top of the production possibility frontier. Therefore, no matter how the number of firms changes, the aggregate output necessarily declines.}

The third row of table 1.2 highlights the implication of the entrepreneurial risks in a
stronger sense. This is the welfare criterion of Garicano et al. [2016]. The result shows that, even if the laissez-faire economy is given the tax advantage, the economy with the SDP still produces more. In this sense, my result is the opposite of Garicano et al. [2016].

The fourth row of table 1.2 shows that the welfare is also higher with the SDP. As discussed in 1.3, the result implies that the SDP increases the aggregate output by non-trivial amount. Quantitatively, the tax revenue

$$T = \phi^{SDP} \mathbb{E} \left[ \left\{ (\tau - 1) w^{SDP} n\left(z, w^{SDP}; \tau, F\right) + F \right\} 1_{n(z, w^{SDP}; \tau, F) > N} \right]$$

(1.18)

amounts to 1.14% of aggregate output in the current setting. The result highlights the first order importance of the channel highlighted in this paper.

The fifth row confirms the decline in wage. As discussed in 1.3, this is an implication of the specific form of the SDP ($\tau, F$). I show in 1.5 that, if the SDP is allowed to be any non-linear function $n \mapsto T(n)$, wage can actually increase. Therefore, even if the government throws away the tax revenue, the SDP can achieve Pareto improvement. In this sense, $w^{SDP} < w^{LF}$ is not necessarily a robust result.

Finally, I discuss these results are robust to the change in the risk aversion parameter $\gamma$. In fact, the break-even point for the output $Y^{SDP} = Y^{LF}$ and welfare $(1 - t) w^{SDP} = w^{LF}$ is $\gamma = .01$. The small risk aversion is not surprising if one notes that the output gain from removing the SDP in the model with no entrepreneurial risks is only .02%. Similarly, the break-even point for $Y^{SDP} - tax = Y^{LF}$ is $\gamma = .7$, still way below the standard parameter value $\gamma = 2$. 

35
1.5 Optimal SDP

The analysis so far has described that an economy with the threshold type SDP \((\tau, F)\) can be better. In other words, an economy with wedges among firms’ marginal products can lead to a better outcome than the one with equal marginal product. A natural question is “what are the optimal wedges implied by SDPs?” To answer this question, I study the optimal non-linear SDP. I also argue that the observed threshold type SDP \((\tau, F)\) shares qualitative features with the optimal SDP.

The optimal SDP is defined as the solution of an optimal taxation problem. Let \(n \mapsto T (n)\) be the tax imposed on the firms that hire \(n\) workers. The government is benevolent and maximizes the welfare

\[
U = \phi \mathbb{E} u (\pi (z, w)) + (1 - \phi) u (w). \tag{1.19}
\]

Although the expression is the same as utilitarian welfare, any Pareto weight leads to the same welfare function, since the full-risk case leads to a homogeneous agents model.

The government faces the market equilibrium constraints and the budget constraint. The market equilibrium constraints consist of the two equations (1.8) and the firm’s profit maximization problem

\[
[\pi (z, n), \pi (z, n)] = \max_{n \geq 0} zf (n) - wn - T (n). \tag{1.20}
\]

Note that the threshold SDP \((\tau, F)\) is a special case where \(T (n) = \{ (\tau - 1) wn + F \} 1_{n > N}\).
The budget constraint takes the following form

\[ \phi \mathbb{E} T(n(z,w)) = T \]  

where \( T \geq 0 \) is exogenously given. There are two reasons for which I do not use the balanced budget constraint. First, by taking the same amount of resources \( T \) as the quantification exercise in section 1.4, I can make a fair comparison between the estimated SDP and the optimal SDP. Second, if \( T = 0 \) is imposed, some firms have to be mechanically subsidized, \( T(n) < 0 \) for some \( n \). By relaxing the constraint \( T > 0 \), I can see whether the government willingly subsidizes some firms.

Formally, the optimal taxation problem can be formulated as follows.

**Definition 1.4.** Given the budget \( T \geq 0 \), the government chooses \((w, \phi, \pi, n, T)\) to maximizes wage \( w \) subject to (1.8), (1.20), and (1.21).

The optimal SDP problem has a similar structure as Mirrlees [1971] except that it has general equilibrium wage. I reformulate it as a mechanism design problem and solve the problem within the set of differentiable functions.

**Proposition 1.4.** The solution to the optimal SDP problem can be obtained as the solution
to

\[
\max_{w, \{n(z), \pi(z)\}_z} w \ \text{s.t.} \quad \begin{cases}
\mathbb{E}u(\pi(z)) - u(w) = 0 \\
\pi'(z) = f(n(z)) \\
n'(z) \geq 0 \\
\mathbb{E}[zf(n(z)) - wn(z) - \pi(z)] - T(1 + \mathbb{E}n(z)) = 0
\end{cases}.
\] (1.22)

\textbf{Proof.} See Appendix A.1.3. \hfill \square

The first constraint is the indifference condition. The second is the envelope condition of the truth-telling constraint, and the third is the monotonicity constraint that corresponds to the second order condition. The last constraint is the government budget constraint.

I can obtain some analytical results about the optimal SDP shared by the observed threshold type SDP. Let \((w^o, \phi^o, n^o, \pi^o, T^o)\) be the solution to the optimal SDP problem.

\textbf{Proposition 1.5.} Suppose that the fundamentals \((u, f, G_z)\) satisfy the assumptions (A), there are risks \(V(z) > 0\), and the solution does not have bunching

\[
\frac{d}{dz} n^o(z) > 0, \quad \forall z \in (z_{\min}, z_{\max}).
\]

Then, the optimal SDP distorts medium-sized firms, i.e., for all \(z \in (z_{\min}, z_{\max}),\)

\[
w^o = z_{\min} f'(n^o(z_{\min})) = z_{\max} f'(n^o(z_{\max})) < z f'(n^o(z)).
\]

For small budget \(T \geq 0\), the optimal SDP increases the number of firms \(\phi^o > \phi^{LF}.\)
Figure 1.7: Marginal product as the measure of distortion and the optimal SDP as the percentage of the before-tax profit. The wage in the economy with the estimated threshold SDP is normalized to be $w_{SDP} = 1$.

**Proof.** See Appendix A.1.4.

As discussed in section 1.3, the threshold type SDP increases the number of firms. Moreover, as Fig.1.5 suggests, the bunching firms are most distorted in the sense that their marginal products are higher than other firms.

To see this point more clearly, I show the numerical solutions in Fig.1.7. The numerical solution is based on the same parameterization as section 1.4, i.e., (1.14) and (1.15), the same parameter values as in table 1.1, and the same tax revenue $T$ as (1.18). One can see that the marginal products are highest in the middle for the threshold type SDP. In particular, the firms that are indifferent between choosing $n = N$ and $n > N$ are most distorted downward.

One thing that can be seen from the left panel is that the wage is higher under the optimal SDP than the threshold SDP by more than 30%. Therefore, table 1.2 implies that the wage level under the SDP is higher than the laissez-faire economy. This is because the
optimal SDP subsidizes small firms as shown in the right figure, and therefore the welfare is higher by the insurance effect.

\[ u \left( w^{SDP} \right) = \mathbb{E}u \left( \pi^{SDP} (z, w) \right) < \mathbb{E}u \left( \pi^{o} (z) \right) = u (w^{o}) . \]

In reality, there are various subsidies available for small firms, so it is not appropriate to say the SDP in reality is not optimal. In fact, the result \( w^{SDP} < w^{LF} < w^{o} \) comes from the parsimonious modeling of the SDP \( (\tau, F) \) that abstract from small firms subsidies. In this paper, I have followed exactly the same formulation as Garicano et al. [2016], but the result can change once the SDP incorporates subsidies for small firms. In this sense, one can actually interpret the quantification exercise in section 1.4 as a strong result saying that, even when the SDP does not subsidize small firms, it could enhance efficiency.

Another thing that can be seen from the left panel of Fig.1.7 is that the optimal SDP increases the number of firms. This is because higher marginal product implies lower employment, which then implies more entrepreneurs in equilibrium. This point can be seen more directly in the left panel of Fig.1.8. One can see that firms with every level of productivity hire fewer workers under the optimal SDP. Similarly, the right panel shows that firms with every level of productivity make lower before-tax profits under the SDP. However, the ex-ante welfare is higher since the optimal SDP insures the bad states of low productivity.

In summary, the main lesson from the study of the optimal SDP is that the wedge among marginal products does not necessarily call for policy intervention to remove the root causes. Rather, it might be better to distort the firm’s production so that the economy features more firm creation and less risky entrepreneurship.
1.6 Financial frictions

The analysis so far has shown that the insufficient firm creation due to entrepreneurial risks might justify SDPs. A natural question is whether this is always the case. I argue the answer is negative by showing that, under severe financial frictions, removing SDPs might improve efficiency. As a policy implication, whether to remove SDPs depends on the level of financial development.

There are three cases where SDPs should be removed. The first case is when entrepreneurial risks are not important. If individual agents obtain a precise signal $s_i$ about their entrepreneurial productivity $z_i$, SDPs are harmful as shown by Garicano et al. [2016] and Gourio and Roys [2014]. The second case is when SDPs are so radical that the efficiency loss dominates the gain from more efficient firm creation. In French data, this is not the case, but data from other countries might find a large estimate of SDP $(\tau, F)$. Both of these cases
Figure 1.9: Aggregate output $Y(\phi)$ when market generates too many firms. Points $A$, $B$, and $C$ are the market equilibrium without SDPs, the market equilibrium with SDPs and the unconstrained efficient allocation.

rely on the idea that SDPs should be removed if the cost is larger than the benefit.

The third case is when there are too many firms in the laissez-faire economy. This case is different from the previous two because removing SDPs has no cost. Fig. 1.9 illustrates this situation. The SDPs that increase the number of firms not only distort production by creating wedges among firms but also worsen the market failure of firm creation by adding more firms to the economy with already excessive firms. Thus, removing SDPs enhances efficiency through two channels.

1.6.1 Financial frictions

One situation that leads to excessive firm creation is severe financial frictions that prevent firms from expanding their employment to their profit-maximizing sizes. I show formally that, under severe financial frictions, the laissez-faire economy creates more firms than the
To illustrate this point parsimoniously, I introduce a simple form of financial friction to the firm’s profit maximization problem as follows.

\[
\pi(z, w) = \max_n zf(n) - wn \text{ s.t. } wn \leq \lambda zf(n)
\]  

(1.23)

where \( \lambda^{-1} > 0 \) represents the severity of the financial friction. If the financial frictions are not severe \( \lambda > \alpha \), the constraint does not bind. Yet as the financial constraint becomes severer, i.e., \( \lambda \) gets close to 0, the constraint becomes more binding. Such form of financial constraint has been used in the literature. A recent example includes Bigio and LaO [2016].

One way to micro-found the constraint is limited enforcement. Each entrepreneur owes the workers their salaries. However, due to the limited enforcement, the entrepreneur can divert \( 1 - \lambda \) fraction of the sales and run without paying the salaries. In this case, workers are not willing to work for the firm if their salaries exceed \( \lambda \) percent of the sales. One can also attribute \( \lambda \) to institutional immaturity or other non-financial obstacles. After all, the formulation is a reduced-form, so I call it financial friction for simplicity.

With this financial constraint, I conduct the efficiency analyses using the full-risk model. Specifically, let \((w^{LF}, \phi^{LF})\) be the equilibrium in the laissez-faire economy characterized by (1.8) and (1.23), \( \phi^P \) be the planner’s solution that solves (1.10) with uninformative signal \( s \perp z \). The following proposition states that under severe financial friction \( \lambda \approx 0 \), the laissez-faire economy generates too many firms, so that the SDPs that increase the number of firms exacerbate human resource misallocation.

**Proposition 1.6.** Fix CRRA utility, CD production function \((u, f)\) and arbitrary risk dis-
tribution $G_z$ with bounded support. Under the severe financial friction, i.e., for small $\lambda \geq 0$, the laissez-faire economy generates excessive number of firms

$$\phi^P < \phi^{LF} \leq 1 = \lim_{\lambda \to 0} \phi^{LF}.$$ 

Proof. See Appendix A.1.5.

The proposition imposes particular functional form assumptions for analytical cleanliness, but they can be relaxed at the cost of heavier notations. The inefficiency result can be strengthened to constrained inefficiency. As is the case for proposition 1.3, I relegate the result to Appendix A.1.5.

To understand the intuition behind the result, recall that the planner is free from financial frictions. As a result, the efficient allocation is invariant to the severity of financial frictions $\lambda$. Therefore, the inefficiency follows if the number of firms in market equilibrium can get arbitrarily close to 1. The intuition of severer financial frictions leading to more firms in the market equilibrium comes from the firm side. If the financial frictions are severe $\lambda \to 0$, due to the cash shortage or limited enforcement, firms cannot expand their employment. In the extreme case $\lambda = 0$, no firm can hire workers. Since the labor demand from firms is limited, those who are not employed have to do their own businesses, leading to an increase in the equilibrium number of firms $\phi \to 1$. Thus, under severe financial friction $\lambda \approx 0$, the number of firms in the laissez-faire economy $\phi^{LF}$ exceeds that of the efficient allocation $\phi^P$.

The analysis of the financial friction implies that whether SDPs should be removed or not depends on the level of financial development. For instance, in developing countries with immature financial infrastructure, SDPs should be removed and resources should be devoted
to relaxing financial frictions, even if firm creation is risky.

This implication might be surprising if one considers that entrepreneurial risks and financial frictions are often cited as two big obstacles to entrepreneurship. In fact, proposition 1.3 and 1.6 suggests that the two frictions cause market failure in the opposite directions, i.e., entrepreneurial risks cause the market to create too few firms while financial frictions cause too many. There are two ways to reconcile the conventional wisdom and my result.

The first one is about the modeling. If one models the financial friction as the fixed cost necessary to start a firm, it does suppress firm creation. However, as Quadrini [2008] discusses, people can lower the fixed cost by choosing to start small. In this sense, the bigger financial friction that entrepreneurs face is the funds to expand, which is exactly what Eq.(1.23) is meant to capture.

The second way is to recognize that the conventional wisdom is based on partial equilibrium, and is not suitable to understand the impact of the financial frictions on the general equilibrium. In partial equilibrium, financial friction reduces profit so people do not want to become entrepreneurs. In general equilibrium, however, since firms cannot expand employment due to the financial friction, those unemployed have to do business themselves. The wage declines to the level consistent with such scenario, rationalizing the existence of many more small firms.

The positive correlation between the severity of financial frictions and firm creation is actually consistent with data. Fig.1.10 plots entrepreneurship activity against the severity of financial frictions. The entrepreneurship is measured by the percentage of the 18 – 65 population who is either a nascent entrepreneur or owner-manager of a new business, taken from Global Entrepreneurship Monitor for 2001 – 2016. The severity of financial frictions is
Entrepreneurship is positively correlated with financial frictions. $x$ axis represents the collateral value required for 100 unit of loans. $y$ axis represents the percentage of the 18 – 65 population who is an either nascent entrepreneur or owner-managers of a new business. The plot pools all available countries and years 2001 – 2016. Sources: World Bank, Enterprise Survey. Global Entrepreneurship Monitor.

measured by the value of collateral needed for 100 unit of loans, taken from the World Bank Enterprise Survey.

One can see that more severe financial frictions are associated with higher entrepreneurial activities. Fig.1.11 presents another data consistent with the analysis. It is based on survey data, asking country experts whether a country has enough entrepreneurship financing. Although the data are subjective, it contains more samples. One can see more availability of financing is associated with lower entrepreneurial activities.

What these figures do not show is the efficiency of the market equilibrium. In other words, they suggest financial frictions are associated with more firm creation, but do not provide information whether the increase is too much or not. In this sense, proposition 1.6 provides a normative insight to the empirical regularities.
Figure 1.11: Entrepreneurship is negatively correlated with the availability of financing. $x$ axis is the survey responses about the availability of financial resources - equity and debt - for SMEs including grants and subsidies. $y$ axis is the percentage of the 18 – 65 population who is an either nascent entrepreneur or owner-managers of a new business. Country year pooled data. Sources: World Bank, Enterprise Survey. Global Entrepreneurship Monitor.

As a caveat, financial frictions are not the entire picture. As one can see from Fig.1.11, there is a difference between high- and low-income countries unexplained by the financial frictions. Indeed, as Gollin [2008], Seshadri and Roys [2014], Poschke [2017] discuss, various explanations can be consistent with the negative correlation between GDP per capita and entrepreneurship. That being said, since the financial friction $\lambda$ in section 1.6.1 is a reduced form formulation, it can also be interpreted as anything that prevents firm’s expansion. In this sense, the unexplained variation does not necessarily limit the scope of the insights.
1.7 Capital accumulation

So far, I have developed analyses in static environments. In this section, I extend the analysis to a tractable dynamic environment with inter-temporal optimization. It is shown that risks justify SDPs even when all agents are risk-neutral. Thus, capital accumulation and consumption smoothing motive further justify SDPs.

To study the implication of inter-temporal trade-off in a parsimonious and tractable dynamic model, I extend the laissez-faire economy by introducing capital accumulation a la Krebs [2003] and Kreps-Porteus-Epstein-Zin preference. (Epstein and Zin [1989], Kreps and Porteus [1978]) The “all-purpose” good a la Krebs [2003] gives dynamic tractability by invoking a similar capital accumulation structure as Brock and Mirman [1972]. The Kreps-Porteus-Epstein-Zin preference isolates the inter-temporal preference from risk aversion so that I can highlight the implications specific to dynamic environments.

1.7.1 Equilibrium

This section defines the equilibrium of the laissez-faire economy in a tractable dynamic environment. I derive the closed-form solution of the equilibrium objects and discuss the similarity to the static model.

The setup of the dynamic economy is a direct extension of the static laissez-faire economy. There is a continuum of agents facing risky occupational choice and dynamic consumption maximization problem. The economy has no aggregate uncertainty.

At the beginning of each period, each agent $i \in [0, 1]$ is characterized by the $(a_{it-1}, j_{it-1}, z_{it})$ consisting of asset $a_{it-1} > 0$, occupation $j_{it-1} \in \{E, W\}$ where $E$ denotes entrepreneurs and
W workers, and productivity $z_{it} > 0$. The distribution of the state at $t$ is denoted by $S_t(a_{it-1}, j_{it-1}, z_{it})$ and the initial distribution $S_0$ is exogenously given.

I make three comments about the state. First, the asset $a_{it}$ is an “all-purpose good” because it can be used as a consumption good, physical capital and human capital. Second, for simplicity I assume the productivity $\{z_{it}\}_{i,t}$ is i.i.d. over $i$ and $t$, distributed according to $G_z$. I assume i.i.d. structure for the simplicity, but it is straightforward to extend it to persistent process that corresponds to the model with the intermediate signal in the static environments. Third, the convention of the time subscript is based on the measurability. In other words, the asset and the occupation are pre-determined at $t − 1$ before observing the productivity $z_{it}$. Therefore, in principle, both the saving and occupation decisions involve risk takings. However, as will be discussed later, the log inter-temporal preference makes sure the only decision affected by entrepreneurial risks is the occupation choice.

Within each period $t$, agents work first, and then make consumption/saving and occupation decisions. For the working part, each agent transforms the all-purpose asset $a_{it-1}$ into productive capital $f_I(a_{it-1})$ to earn income. Each entrepreneur $i \in E_t$ uses the capital $k_{it-1} = f_I(a_{it-1})$ to run her firm of productivity $z_{it}$. In particular, given wage $w_t$, the entrepreneur $i$ solves the profit-maximization problem to obtain $\pi_{it} = \pi(z_{it}, w_t, k_{it-1})$ by hiring $n_{it} = n(z_{it}, w_t, k_{it-1})$ employees

$$[\pi(z, w, k), n(z, w, k)] = \max_{n \geq 0} zf(k, n) - wn. \quad (1.24)$$

Each worker $i \in W_t$ sells his human capital $h_{it-1} = f_I(a_{it-1})$ to earn labor income $w_t h_{it-1}$.

After obtaining income, each agent $i$ makes consumption-saving decision and chooses
whether to become an entrepreneur or a worker next period. The consumption-saving decision \((c_{it}, a_{it})\) has to be chosen from the budget set

\[
B_t(a, E, z) = \{ (c, a') : k = f_I(a), c + a' = \pi(z, w_t, k) \}
\]

\[
B_t(a, W, z) = \{ (c, a') : h = f_I(a), c + a' = w_t h \}.
\] (1.25)

The occupation choice allows mixed strategy \(e_{it} \in [0, 1]\) as in the static case, i.e., if an agent chooses \(e_{it}\), she becomes an entrepreneur \(j_{it} = E\) with probability \(e_{it}\). Each agent \(i\) chooses \((c_{it}, a_{it}, e_{it}) \in B_t(a_{it-1}, j_{it-1}, z_{it}) \times [0, 1]\) for all states to maximize the recursive utility

\[
V_{it} = u_I^{-1} \left[ (1 - \beta) u_I(c_{it}) + \beta u_I \left( \mathbb{E} u \left( V_{it+1} \right) \right) \right]
\] (1.26)

where \(u\) and \(u_I\) are intra- and inter-temporal utility functions, and the expectation \(\mathbb{E}\) is taken with respect to both productivity \(z_{it} > 0\) and the mixed strategy \(e_{it} \in [0, 1]\).

Finally, the labor market has to clear, i.e., labor demand from entrepreneurs equal to human capital supply from workers.

\[
\int e_{it-1} n_{it} di = \int (1 - e_{it-1}) h_{it-1} di,
\] (1.27)

The following definition summarizes the equilibrium concepts.

**Definition 1.5.** Fix the fundamentals \(\mathcal{E} = (\beta, u, u_I, f, f_I, G_z, S_0)\). The set of individual choices and wage \(\{V_{it}, a_{it}, c_{it}, e_{it}, h_{it}, k_{it}, \pi_{it}, n_{it}, w_t\}_{i,t}\) constitutes an equilibrium if the following conditions are satisfied for all \(t \geq 0\).
1. At each \( t \), given \( w_t \), the profit and the employment decision \( \{ \pi_{it}, n_{it} \} \) by the entrepreneur with productivity \( z_{it} \) solves the profit maximization problem (1.24).

2. Given \( \{ w_t \}_t \), the value and policies \( \{ V_{it}, a_{it}, c_{it}, e_{it} \} \) maximize individual welfare (1.26)
   subject to \( (c_{it}, a_{it}, e_{it}) \in B_t (a_{it-1}, j_{it-1}, z_{it}) \times [0, 1] \).

3. At each period \( t \), labor market clears (1.27).

The equilibrium number of firms at \( t \) is pre-determined at \( t - 1 \) and can be written as

\[
\phi_{t-1} = \int c_{it-1} di.
\]

This is a dynamic general equilibrium model with heterogeneous agents and incomplete markets, so in general, a closed form solution is not available. However, the following parameterizations allow us to study the efficiency of firm creation analytically

\[
u (c) = \frac{c^{1-\gamma} - 1}{1 - \gamma}, \quad u_I (c) = \ln c, \quad f_I (k, n) = k^{1-\alpha} n^\alpha, \quad f_I (a) = a^\theta \quad (1.28)
\]

where \( \gamma > 0, \quad \alpha \in (0, 1) \) and \( \theta \in (0, 1) \). The intra-temporal utility \( u \) remains the same as the static model. The logarithmic inter-temporal utility \( u_I \) separates the saving decision from the uncertainty about the productivity \( z_{it} \). This is because the optimal saving is a constant fraction of the income and is independent of the potentially risky returns to saving. In this way, I can make sure the only risky decision is occupation choice, so that the result is directly comparable with the static environment. The production function is the standard Cobb-Douglas form with complementarity between the entrepreneurial capital and the workers’ human capital. As a result, the profit becomes a linear function of entrepreneurial capital.
\[ \pi(z, w, k) = r(z, w) k, \] and therefore can be interpreted as the return to entrepreneurial capital. The investment function features decreasing returns to scale, which ensures that there is a stationary asset distribution.

Given these parameterizations, the market equilibrium takes the following form as derived in Appendix A.1.6

\[
\begin{align*}
\phi_t^{LF} &= \frac{1 - \alpha}{1 - \alpha + \alpha \mathbb{E}_z \frac{1}{1 - \alpha} \left( \mathbb{E}_z \frac{(1 - \gamma)(1 - \beta)}{(1 - \alpha)(1 - \beta \theta)} - \frac{1 - \beta \theta}{1 - \gamma(1 - \beta)} \right)}, \\

w_t^{LF} &= \left( \mathbb{E}_z \frac{(1 - \gamma)(1 - \beta)}{(1 - \alpha)(1 - \beta \theta)} \right) \alpha^\alpha (1 - \alpha)^{1 - \alpha},
\end{align*}
\]

One can see that the risk aversion \(1 - \gamma\) is always multiplied by \(\frac{1 - \beta}{1 - \beta \theta} > 1\). This additional term comes from the inter-temporal decision and makes the agents effectively more risk-averse. In fact, even when agents are risk-neutral \(\gamma = 0\), the number of firms \(\phi_t^{LF}\) is smaller than \(1 - \alpha\), the efficient number of firms in the static environment. As shown in the next section 1.7.2, \(1 - \alpha\) is also the efficient number of firms in the dynamic environment.

1.7.2 Efficiency analysis

This section presents the efficiency analysis of the dynamic model. The direction of the market failure remains the same as in the static economy but the severity is greater due to the consumption-smoothing motive. A novel insight is that the market failure is more severe in an economy with more patient agents.

The planner maximizes the welfare by choosing resource allocations at each state \(\omega_t \in \Omega_t = \left( [z_{min}, z_{max}]^{(0,1)} \right)^t\) facing the same technology and information constraints as the indi-
individual agent. Specifically, given the Pareto weight function \( i \mapsto \Lambda_i \), the planner maximizes the welfare \( \int V_{it}(\omega) d\Lambda_i \) where \( V_{it}(\omega) \) satisfies

\[
V_{it}(\omega) = u_{i}^{-1} \left[ (1 - \beta) u_I(c_{it}(\omega_t)) + \beta u_I \left( \mathbb{E}_{\Omega_{t+1}|\omega_t} u(V_{it+1}) \right) \right]
\]

(1.30)

and the expectation \( \mathbb{E}_{\Omega_{t+1}|\omega_t} \) is taken over the subset of state space \( \Omega_{t+1} \) accessible from \( \omega_t \). In the following, I omit the notation of the state \( \omega \) and related measure theoretic treatment, since the problem reduces to the aggregate production maximization in an economy with no aggregate uncertainty.

The planner’s decision can be described recursively. At each period \( t \), the planner takes as given the set of entrepreneurs \( E_{t-1} \subset [0, 1] \) and workers \( W_{t-1} = [0, 1] \setminus E_{t-1} \) chosen at period \( t - 1 \), and their saving for capital, \( \{a_{it-1}\}_i \). The capital generates entrepreneurial capital \( k_{it-1} = f_I(a_{it-1}) \) if stored by an entrepreneur \( i \in E_{t-1} \) and human capital \( h_{it-1} = f_I(a_{it-1}) \) if stored by a worker \( i \in W_{t-1} \). The human capital is then reallocated by the planner to make production

\[
Y_t = \int_{i \in E_{t-1}} z_{it} f(k_{it-1}, n_{it}) \, di, \quad \int_{i \in E_{t-1}} n_{it} \, di = \int_{i \in W_{t-1}} h_{it-1} \, di.
\]

(1.31)

After the production, the planner chooses the occupation \( E_t, W_t \) for period \( t + 1 \) and divides the aggregate output into consumption today and asset saved for tomorrow

\[
\int c_{it} \, di + \int a_{it} \, di = Y_t.
\]

(1.32)

The key informational assumption is that when the planner chooses the asset \( a_{it} \), the pro-
ductivity tomorrow \( z_{it+1} \) is not observable

\[
a_{it} \perp z_{it+1}. \tag{1.33}
\]

In summary, the efficient allocation can be defined as follows.

**Definition 1.6.** Fix the fundamentals \( \mathcal{E} \). The set of individual values, consumption-saving, occupation, and production decisions and the aggregate output \( \{ V_{it}, c_{it}, a_{it}, E_t, W_t, k_{it}, h_{it}, n_{it}, Y_t \} \) is efficient if it maximizes (1.30) subject to (1.32), (1.33) and (1.31). The number of firms in the planner’s solution is defined as

\[
\phi_{t-1}^P = \int_{i \in E_{t-1}} di.
\]

It is immediate to see that the efficient allocation maximizes the aggregate output \( Y_t \) at each period given the asset \( \{a_{it-1}\} \). Since the planner can control all the allocations, as in the static case, the planner’s problem can be separated into production and allocation of consumption.

The following proposition states that the direction of the market failure remains the same as the static case, but the market failure gets worse due to the inter-temporal consumption smoothing motive.

**Proposition 1.7.** The laissez-faire economy generates an insufficient number of firms if and only if there are entrepreneurial risks.

\[
\phi_{t}^{LF} < \phi_{t}^{P} = 1 - \alpha \iff V(z) > 0.
\]
This is true even when all agents are risk neutral \( \gamma = 0 \).

\textit{Proof.} See Appendix A.1.6.

The inefficiency result can be extended to constrained inefficiency as discussed in Appendix A.6.3. I make two observations about the statement. First, note that the planner’s solution \( \phi^P_t = 1 - \alpha \) is identical to the static model, making the static and dynamic models directly comparable. Second, the last line presents a new insight specific to the dynamic environment. It states that the market failure happens even when all agents are risks-neutral. This is not the case in the static environment. If agents are risk-neutral, both the laissez-faire and planner choose \( \phi = 1 - \alpha \) in the static environment, making Pareto improvement impossible.

The intuition is based on the consumption-smoothing motive. Since agents want to smooth consumption, they want to avoid asset fluctuation. To avoid the asset fluctuation, one can choose the salary job with a safer return. Therefore, the number of firms is smaller than the efficient level.

The key observation is that consumption-smoothing motive works similarly as risk aversion. This is not surprising if one realizes that the intra-temporal and inter-temporal preferences both discount the value of fluctuations across a set of states. Mathematically, they both represent the force of Jensen’s inequality over the states called in different ways.

However, the observation does provide a non-trivial insight about the patience \( \beta \). As can be seen from Eq.(1.29), the more patient agents are \( \beta \nearrow \), the severe the market failure is \( \phi^P - \phi^{LF} \nearrow \). This is because patient agents care more about the fluctuation of the future consumption. In other words, higher patience magnifies consumption-smoothing motive.
Therefore, in high patience economy, removal of SDPs can cause severe human resource misallocation.

1.8 Conclusion

I have shown that SDPs might enhance efficiency in the presence of entrepreneurial risks. The analysis not only changes the conventional wisdom about SDPs, but also casts a caveat in general to the misallocation measurement exercises based on marginal product equalization. I have also made the analysis comprehensive by showing that severe financial frictions can cause excessive firm creation even if firm creation is risky. This result implies that developing countries with severe financial frictions should remove SDPs. Finally, I have analyzed the implications of dynamic decisions. The analysis not only confirms the robustness of the findings in the static environment but also provides a novel insight into the role of consumption-smoothing motive and patience. These analyses deepen our understanding of SDPs.

I conclude the paper by mentioning three caveats to my analyses. First, there can be many other justifications for SDPs. For instance, small firms might have some externality in R&D and employment. However, these arguments are highly controversial compared to entrepreneurial risks (Biggs [2003], Shane [2008, 2009]). Another justification for SDPs is that it saves the government administrative costs as discussed in Kaplow [2017]. My analysis should be considered a complement to these discussions. Second, I have assumed that individual agent maximizes profit, but the importance of non-pecuniary benefit has also been pointed in the entrepreneurship literature (Hurst and Pugsley [2011], Gordon and
Sarada [2017]). Exploring the implications of behavioral agents that have other objectives than profit maximization is an interesting future topic. Third, I have focused on SDPs that treat small firms preferentially. However, some countries treat large firms preferentially to foster international competitiveness. In this case, SDPs are adopted based on a totally different logic and therefore their analysis needs a different framework. These points are missing in the analysis and should be explored in future research.
Chapter 2

Constrained Efficiency of Competitive Entrepreneurship

Sakai Ando and Misaki Matsumura
2.1 Introduction

It is well known that incomplete markets are generically constrained inefficient (Diamond [1967], Greenwald and Stiglitz [1986], Geanakoplos and Polemarchakis [1986], Geanakoplos et al. [1990]). When the incompleteness comes from uninsurable risks, the constrained planner is known to improve welfare by providing insurance through changing equilibrium prices. Despite the established mathematical argument and its generality, however, few examples are available that are so simple that one can grasp the economic intuition of why incomplete markets create room for the constrained planner’s intervention. Since constrained inefficiency in incomplete markets is not just of theoretical interest but often is used as the justification for policy intervention, an example in which one can follow the precise logic without relying on mathematical derivation is valuable.

In this paper, we use the competitive entrepreneurship model, a descendant of Kanbur [1979b], to illustrate the economic intuition of why the competitive equilibrium leaves room for intervention by the constrained planner. The key feature of the competitive entrepreneurship model is the simple structure of the equilibrium price determination. Specifically, the equilibrium price makes sure the certainty equivalents of all agents are equated. Therefore, when there are no risks, which corresponds to complete markets in the competitive entrepreneurship model, welfare gains from marginal redistribution are canceled out, ending up with no welfare improvement in net. However, when there are risks in entrepreneurship, the constrained planner’s intervention that uses price change to redistribute consumption from risk-free jobs to risky jobs has different impacts on the certainty equivalents of the two groups. Therefore, the constrained planner can find a welfare-improving intervention.
Such explanation differs from saying that there is a pecuniary externality and the constrained planner can improve welfare by internalizing it. The latter offers a verbal summary of the mathematical results, but it does not explain why constrained planner can improve welfare only when the markets are incomplete. The fact that agents’ decisions have pecuniary externalities is true no matter whether markets are complete or not, but the constrained planner can find welfare-improving intervention only when the markets are incomplete. In other words, since the instruments that the constrained planner can use are exactly the same as agents in the market equilibrium, it is not clear why the market cannot internalize the pecuniary externality. To get the economic intuition, therefore, it is necessary to understand what the market does and how the constrained planner improves the welfare only when markets are incomplete.

To highlight the subtlety of the argument, we contrast the competitive entrepreneurship model with Aiyagari and Krebs models. Both of them are known to verify constrained inefficiency if and only if there are uninsurable risks, and have structures simple enough to be explained without heavy notation. The former features consumption-saving decision while the latter illustrates the portfolio choice. By showing these two examples, we argue that the difficulty of intuitively explaining constraint inefficiency is a prevalent problem. We quote the explanations of the intuitions from the Davila et al. [2012] and Toda [2015], and point out what is missing from the intuitions.

Our paper belongs to the literature of the constrained efficiency analysis pioneered by Diamond [1967]. Our result shares the same flavor as the generic constrained inefficiency of incomplete markets. Geanakoplos and Polemarchakis [1986], Greenwald and Stiglitz [1986], Geanakoplos et al. [1990]) However, the determinant of constrained inefficiency in our model
is not the position of endowment vectors but is the existence of entrepreneurial risks. In
addition, by exploiting the simple structure where agents face only two alternatives, we can
clarify the intuition of how exactly incomplete market leads to constrained inefficiency. Other
papers that analyze constrained efficiency by exploiting the simple structure of two choices
include Farhi et al. [2009] that study Diamond and Dybvig [1983], Davila et al. [2012] that
study Aiyagari [1994] and Toda [2015], Gottardi et al. [2016] that study Krebs [2003]. We
contribute to this strand of literature by studying the constrained efficiency of the standard
occupation choice model. (Lucas [1978], Kanbur [1979b,a, 1982]) Davila and Korinek [2017]
study the constrained efficiency in a model with both risks and collateral constraints.

2.2 Two simple examples of constrained inefficiency

In this section, we show two simple models in which the market equilibrium is constrained
inefficient. We argue that the difficulty of understanding the economic intuition comes from
the interpretation of the marginal condition that characterizes the market equilibrium.

2.2.1 Aiyagari model

Following Davila et al. [2012], we illustrate the constrained inefficiency using the two-period
version of Aiyagari [1994].

The economy consists of ex-ante identical households of mass 1 endowed with 1 unit of
asset and a representative firm with constant returns to scale technology. In the first period,
households decide the amount of saving $a$. The saving generates return $r$ in the second period.
Households on average earn labor income $w$ in the second period but face idiosyncratic labor
endowment risks $\sigma$ distributed according to $G$ in the second period satisfying $\mathbb{E}\sigma = 1$ and

$P(\sigma > 0) = 1$. Formally, households solve

$$\max_a u(1 - a) + \mathbb{E}u(ra + w\sigma),$$

(2.1)

where the utility function $u$ is increasing, concave and twice differentiable. The representative firm uses capital and labor to produce.

$$\max_{K,L} f(K, L) - rK - wL$$

(2.2)

Since the production technology is constant returns to scale, we do not have to specify dividend payment. Finally, markets for saving/capital and labor clear.

$$a = K, \quad 1 = L.$$  

(2.3)

The equilibrium is defined as follows.

**Definition 2.1.** Fix $(u, f, \sigma)$. $(a, r, w)$ is a market equilibrium if the following three conditions are satisfied.

(i) Given $(r, w)$, $a$ solves equation (2.1).

(ii) Given $(r, w)$, $(a, 1)$ solves equation (2.2).

(iii) Markets clear (2.3).

The market equilibrium saving $a^m$ can be characterized by a single equation. To see this,
note that prices can be solved in closed form from (2.2)

\[(r(a), w(a)) := (f_K(a, 1), f_L(a, 1)).\]

Substituting these prices into the household’s first order condition gives

\[u'(1 - a^m) = r(a^m) E u'(r(a^m) a^m + w(a^m) \sigma). \tag{2.4}\]

One can interpret Eq. (2.4) as equating marginal benefit and cost, but it is not clear whether there is room for efficiency improvement or not.

The constrained planner examines the efficiency by choosing savings on behalf of households, but taking into account all the general equilibrium effects.

**Definition 2.2.** Fix \((u, f, \sigma)\). \(a^{cp}\) is constrained efficient if

\[a^{cp} = \arg \max_a u(1 - a) + E u(r(a) a + w(a) \sigma).\]

For the constrained planner’s problem to be concave, we need to impose conditions on the third derivative of the production function \(f''_m\). This is because the prices \((r(a), w(a))\) are derived from the first-order conditions. Since there are no standard assumptions on the third derivative of production functions, we simply assume the production function is Cobb-Douglas in the following analyses.

As shown in Davila et al. [2012], the constrained planner chooses less saving \(a^{cp} < a^m\), so the households save too much. Since Davila et al. [2012] solves the special case where \(\sigma\)
takes two values, we restate the result in a general form. The proof that follows shows the
typical argument of the constrained inefficiency in the literature.

**Proposition 2.1.** *If the production function* \( f \) *takes the Cobb-Douglas (CD) form, the com-
petitive market creates excessive saving* \( a^p < a^m \) *if and only if there are risks* \( V(\sigma) \neq 0. \)

*Proof.* The proof proceeds in two steps. First, we show that the planner’s problem is concave.
Define the objective function as

\[
U(a) := u(1 - a) + \mathbb{E}u(r(a) a + w(a) \sigma).
\]

Then, the first and second order derivatives can be written as

\[
U'(a) = -u'(1 - a) + \mathbb{E}[u'(r(a) a + w(a) \sigma)(r(a) + r'(a) a + w'(a) \sigma)] \quad (2.5)
\]

\[
U''(a) = u''(1 - a) + \mathbb{E}\left[u''(r(a) a + w(a) \sigma)(r(a) + r'(a) a + w'(a) \sigma)^2\right]
+ \mathbb{E}[u'(r(a) a + w(a) \sigma)(r''(a) a + 2r'(a) + w''(a) \sigma)]
\]

Now, recall the definition \((r(a), w(a)) := (f_K(a, 1), f_L(a, 1))\). If \( f(K, L) = K^\alpha L^{1-\alpha} \), then
\[
w''(a) = f_{LKK}(a, 1) = -\alpha(1 - \alpha)^2 a^{\alpha-2} < 0\]
and
\[
r''(a) a + 2r'(a) = f_{KKK}(a, 1) a + 2f_{KK}(a, 1) = -\alpha^2 (1 - \alpha) a^{\alpha-2} < 0.
\]

Hence, the problem \( U(a) \) is strictly concave.

Second, we show that the planner’s objective function is strictly decreasing when evalu-
ated at the market equilibrium. To see this, evaluate the constrained planner’s first order derivative (2.5) at the market equilibrium. Then, from equation (2.4),

\[ U'(a^m) = E \left[ u'(r(a^m)a^m + w(a^m)\sigma)(r'(a^m)a^m + w'(a^m)\sigma) \right]. \]

Since \( f \) is homogeneous of degree 1, differentiating \( f_K(K, L)K + f_L(K, L)L = f(K, L) \) with respect to \( K \) gives

\[ f_{KK}(a, 1)a + f_{KL}(a, 1) = 0. \]

Hence, the first order condition evaluated at the market equilibrium

\[ U'(a^m) = E \left[ u'(r(a^m)a^m + w(a^m)\sigma)r'(a^m)a^m(1 - \sigma) \right] \]

\[ = \text{Cov}(u'(r(a^m)a^m + w(a^m)\sigma), f_{KK}(a^m, 1)a^m(1 - \sigma)) \]

is negative iff \( V(\sigma) > 0 \). Hence, \( a^m > a^\text{sp} \) iff \( V(\sigma) > 0 \).

As can be seen from the proof, the key step is that the first-order condition evaluated at market equilibrium \( a^m \), \( U'(a^m) \), combined with (2.4), reduces to the pecuniary externality term.

\[ U'(a^m) = E \left[ u'(r(a^m)a^m + w(a^m)\sigma)(r'(a^m)a^m + w'(a^m)\sigma) \right]. \]

This term can be called pecuniary externality term because it represents the welfare impact of increasing saving through prices \((r'(a), w'(a))\). This pecuniary term is 0 if and only if there are no risks \( V(\sigma) = 0 \), or the market is complete. Davila et al. [2012] explains the constrained inefficiency as follows.
The intuitive reason for the overaccumulation of capital is as follows. More capital savings raises wages and lowers rental rates. The only source of market failure in this economy is the incomplete insurance. A small decrease in $a$ from the equilibrium level thus lowers $w$ and raises $r$, thereby scaling down the part of the consumer’s income that is stochastic and scaling up the part that is deterministic: the amount of risk the consumer is exposed to is now smaller. Given that there is no direct insurance for this risk, this amounts to an improvement. The “distortion” on the agents’ savings by moving savings away from the competitive equilibrium level for given prices is of a second-order magnitude, and thus the manipulation of prices so as to lower the de facto risk dominates.

The problem of the intuition is that the fact that the planner’s intervention reduces risk exposure does not automatically imply improvement. The constrained planner can further reduce risk by decreasing $a^{cp}$ to $a^{cp} - \epsilon$, but $a^{cp} - \epsilon$ is not an improvement from $a^{cp}$. In other words, the key step of the logic is to understand why market equilibrium is not optimal, but the above explanation does not address it. In order to understand why market equilibrium is not optimal, we need a model in which one can see clearly what the market does.

### 2.2.2 Krebs model


The economy consists of ex-ante identical households of mass 1 endowed with 1 unit of
the all-purpose good and a representative firm with constant returns to scale technology. All-purpose good can become either physical capital or human capital. Households decide how much to invest in physical capital, which generates return $r$, and how much in human capital, which generates return $w$. As in Aiyagari [1994], households face idiosyncratic labor risks $\sigma$ distributed according to the distribution function $G$ with $\mathbb{E}\sigma = 1$ and $P(\sigma > 0) = 1$. Formally, households solve

$$\max_{k,l} \mathbb{E} u(rk + \sigma wl) \ s.t. \ k + h = 1. \quad (2.6)$$

The representative firm uses physical capital and human labor to produce.

$$\max_{K,L} f(K, L) - rK - wL \quad (2.7)$$

Since the production technology is constant returns to scale, we do not have to specify dividend payment. Finally, markets for both types of capital clear

$$k = K, \ l = L. \quad (2.8)$$

**Definition 2.3.** Fix $(u, f, G)$. $(k, l, K, L, r, w)$ is a market equilibrium if the following three conditions are satisfied.

(i) Given $(r, w)$, $(k, l)$ solves equation (2.6).

(ii) Given $(r, w)$, $(K, L)$ solves equation (2.7).

(iii) Markets clear (2.8).
The equilibrium prices can be solved in closed form

\[(r(k), w(k)) = (f_K(k, 1 - k), f_L(k, 1 - k)).\]

By substituting prices out from the households’ optimization condition, the market equilibrium saving level \(k^m\) can be characterized by a single equation.

\[
\mathbb{E}[u'(r(k^m)k^m + \sigma w(k^m)(1 - k^m))(r(k^m) - \sigma w(k^m))] = 0. \tag{2.9}
\]

The constrained planner chooses savings on behalf of households, but it takes into account all the general equilibrium effects.

**Definition 2.4.** Fix \((u, f, G)\). \(a^{cp}\) is constrained efficient if

\[k^{cp} = \arg \max_k \mathbb{E}u(r(k)k + \sigma w(k)(1 - k)).\]

One can see the constrained inefficiency using a similar argument.

**Proposition 2.2.** If the production function \(f\) takes the Cobb-Douglas form, the competitive market creates excessive capital investment \(k^{cp} < k^m\) if and only if there are risks \(V(\sigma) > 0\).

**Proof.** First, the planner’s problem is concave. Define

\[U(k) = \mathbb{E}u(r(k)k + \sigma w(k)(1 - k)).\]
Then, since $f(K, L) = K^\alpha L^{1-\alpha}$,

$$U(k) = \mathbb{E}u\left(\{\alpha + \sigma (1 - \alpha)\} k^\alpha (1 - k)^{1-\alpha}\right).$$

Since $k \mapsto k^\alpha (1 - k)^{1-\alpha}$ is a strictly concave function, so is its strictly concave transformation $u\left(\{\alpha + \sigma (1 - \alpha)\} k^\alpha (1 - k)^{1-\alpha}\right)$, and therefore the weighted sum $U(k)$ is strictly concave.

One can see that the planner’s solution is $k = \alpha$.

$$U'(k) = 0 \iff k = \alpha.$$

If the planner’s FOC is evaluated at market equilibrium $k^m$, we have

$$U'(k^m) = \mathbb{E}\left[u'\left(r\left(k^m\right) k^m + \sigma w\left(k^m\right) (1 - k^m)\right) (r'\left(k^m\right) k^m + r\left(k^m\right) (1 - k^m) - \sigma w\left(k^m\right))\right]$$

$$= \mathbb{E}\left[u'\left(r\left(k^m\right) k^m + \sigma w\left(k^m\right) (1 - k^m)\right) (r'\left(k^m\right) k^m + \sigma w'\left(k^m\right) (1 - k^m))\right],$$

where the second equality follows from households’ optimization condition. Note that under the Cobb-Douglas assumption,

$$\mathbb{E}\left[r'(k) k + \sigma w'(k) (1 - k)\right] = \frac{\alpha (1 - \alpha)}{k (1 - k)} k^\alpha (1 - k)^{1-\alpha} \mathbb{E} (\sigma - 1) = 0.$$

Hence, $U'(k^m)$ can be written as the covariance of a decreasing and increasing functions of $\sigma$

$$U'(k^m) = Cov\left(u'\left(r\left(k^m\right) k^m + \sigma w\left(k^m\right) (1 - k^m)\right), \frac{\alpha (1 - \alpha)}{k^m (1 - k^m)} (k^m)^\alpha (1 - k^m)^{1-\alpha} (\sigma - 1)\right).$$
This term is strictly negative iff $\sigma$ has variance. Hence, $k^{cp} < k^m$ iff $V(\sigma) > 0$.

Toda [2015] explains the intuition as follows.

The intuition for the generic constraint inefficiency result is straightforward. ... the return on an individual’s portfolio depends on the portfolio choice of other agents through the effect on rental rates. In essence, there is a ‘portfolio externality’, which makes the economy inefficient.

The problem of the intuition is that the portfolio externality exists even when there are no risks $V(\sigma) = 0$. Therefore, the portfolio externality or pecuniary externality itself does not imply constrained inefficiency. The economic intuition has to address why pecuniary externality matters only when there are risks, and does not when there are no risks. Again, the economic intuition requires understanding why market equilibrium is not efficient, but the structure of the model is not suitable to see what the market does.

### 2.3 Competitive entrepreneurship model

The economic intuition of constrained inefficiency can be obtained in the competitive risky entrepreneurship model. Such model traces back to Kanbur [1979a], but its constrained efficiency result has not been investigated. We show that the model makes the economic intuition of constrained inefficiency transparent because it is easy to see what the market equilibrium does.

There is a continuum of ex-ante identical risk-averse agents of mass 1. Each agent is endowed with 1 unit of labor that can be spent in running a firm or working for a firm. If she
chooses to be an entrepreneur, she observes her entrepreneurial productivity $z_i$ distributed according to $G$, and then given wage $w$, decides the number of employees $n$ to maximize the profit

$$[\pi(z, w), n(z, w)] = \max_n zf(n) - wn. \quad (2.10)$$

The notation indicates $\pi$ and $n$ are the value and policy functions of the maximization problem on the right-hand side. If she chooses to become a worker, she receives wage $w$ independent of her entrepreneurial productivity $z$. Given the payoffs of the two occupations, each agent chooses the occupation that maximizes her welfare

$$\phi = \max_e e \mathbb{E}[u(\pi(z, w))] + (1 - e) u(w) \quad (2.11)$$

where the expectation is taken with respect to the productivity $z$. Note that the choice variable $e$, representing whether to become an entrepreneur, can take continuous values $e \in [0, 1]$. This formulation allows agents to take mixed strategy. We focus on the symmetric equilibrium where all agents take the same strategy $\phi$. Since the total population is 1, $\phi$ also represents the mass of entrepreneurs. Hence, the labor market clearing condition becomes

$$\phi \mathbb{E}n(z, w) = 1 - \phi. \quad (2.12)$$

where the left-hand side is the labor demand and right side denotes the labor supply.

The market equilibrium is defined as follows.

**Definition 2.5.** Fix $(u, f, G)$. $(\pi, n, \phi, w)$ is a market equilibrium if the following three conditions are satisfied.
(i) Given $w$, $(\pi, n)$ solves (2.10).

(ii) Given $w$, $\phi$ solves (2.11).

(iii) Market clears (2.12).

What makes the framework suitable for understanding the constrained inefficiency is that it is easy to see what the market equilibrium does since the optimality condition does not involve derivatives. In particular, the market equilibrium wage $w^{m}$ makes sure the two occupations are indifferent in equilibrium.

$$
\mathbb{E} u (\pi (z, w^{m})) = u (w^{m}).
$$

(2.13)

In other words, the market makes sure the certainty equivalent of the two occupations equal.

We consider the constrained efficiency of the occupation choice. Suppose the planner chooses occupation on behalf of agents, but just as in sections 2.2.1 and 2.2.2, the planner does not intervene in the production decisions. Therefore, the price $w (\phi)$ can be implicitly defined by the market clearing condition (2.12).

**Definition 2.6.** Fix $(u, f, G)$. $\phi^{cp}$ is constrained efficient if

$$
\phi^{cp} = \arg \max_{\phi} \phi \mathbb{E} u (\pi (z, w (\phi))) + (1 - \phi) u (w (\phi)).
$$

Compared to the market equilibrium, the constrained planner is not constrained by the indifference condition (2.13).

A similar mathematical argument as sections 2.2.1 and 2.2.2 leads to the constrained
inefficiency result. To illustrate the economic intuition, we present a heuristic explanation and the corresponding mathematical equations, relegating the formal mathematical proof to Appendix B.1.1.

**Proposition 2.3.** Let $\phi^m$ be the market equilibrium number of entrepreneurs. If $(u, f)$ take the form of constant relative risk aversion (CRRA) and CD, $\phi^m < \phi^{cp}$ if and only if there are risks $V(z) > 0$.

*Proof.* See Appendix B.1.1.

We divide the intuition into two parts, why the market equilibrium is constrained inefficient $\phi^m \neq \phi^{cp}$ and why the direction of market failure points to insufficiency $\phi^m < \phi^{cp}$.

### 2.3.1 Why is market equilibrium constrained inefficient $\phi^m \neq \phi^{cp}$?

The intuition of the constrained inefficiency $\phi^m \neq \phi^{cp}$ can be understood by using certainty equivalent. Note that the certainty equivalent of the workers is the wage itself $c^w = w$ since it is risk-free, while the certainty equivalent of the entrepreneurs $c^E$ is defined by

$$u(c^E) = \mathbb{E}u(\pi(z, w)).$$  \hspace{1cm} (2.14)

As a result of the indifference condition (2.13), $c^E = c^w$, so the marginal utilities evaluated at the certainty equivalents are equated

$$u'(c^E) = u'(c^w).$$  \hspace{1cm} (2.15)

This is true no matter whether there are entrepreneurial risks or not.
What the planner can do is to change the wage $w$ by changing the number of entrepreneurs $\phi$ through market equilibrium. To understand how the planner improves, suppose for simplicity that there are the same number of workers and entrepreneurs $\phi^m = \frac{1}{2}$, and that the aggregate output is fixed $Y(\phi^m)$ so that the constrained planner’s intervention is a pure redistribution. The equilibrium number of entrepreneurs $\phi^m$ can be controlled by varying the production technology $\alpha$, and Appendix B.2 shows that output is indeed constant for a marginal intervention by the constrained planner. Now, suppose the planner increases the wage $w$ marginally by 1 unit from the market equilibrium $w^m$.

If there are no entrepreneurial risks, consumption and certainty equivalents are identical. The fact that there are equal mass of workers and entrepreneurs and output is fixed implies that, when workers’ consumption increases by 1 unit, entrepreneurs’ consumption decreases by exactly the same 1 unit $\Delta c^W = -\Delta c^E = 1$. Since the marginal utilities of the certainty equivalents are equated (2.15), the welfare gain from the workers $u'(c^W) \Delta c^W$ is exactly canceled out by the welfare loss of the entrepreneurs $u'(c^E) \Delta c^E$, ending up with 0 welfare gain in net

$$u'(c^E) \Delta c^E + u'(c^W) \Delta c^W = u'(c^W) (\Delta c^E + \Delta c^W) = 0.$$  

This is why the market is constrained efficient when there are no entrepreneurial risks.

If there are entrepreneurial risks, however, the 1 unit increase of the workers’ consumption $\Delta c^W = 1$ does not necessarily correspond to the 1 unit decrease of the entrepreneurs’ certainty equivalent of consumption $\Delta c^E \neq -1$, since the wage change might affect the distribution of profit $\pi(z, w)$ and therefore affects both the expected profit and its risk premium.
As a result, the utility differences might not be canceled out

\[ u'(c^E) \Delta c^E + u'(c^W) \Delta c^W = u'(c^W) (\Delta c^E + \Delta c^W) \neq 0, \]

so that the planner can find welfare gain out of this redistribution. Note that the difference of the argument from no-risk case is whether the wage change affects the risk premium of the entrepreneurs’ consumption or not.

This heuristic explanation can be backed by the corresponding equations. To see this, note that the first-order condition of the constrained planner evaluated at the market equilibrium is

\[ U'(\phi^m) = \left\{ \phi^m \frac{\partial E u (\pi (z, w^m))}{\partial w} + (1 - \phi^m) \frac{\partial u (w^m)}{\partial w} \right\} w'(\phi^m). \tag{2.16} \]

In terms of the economics, Eq. (2.16) corresponds to the thought experiment of the marginal wage increase that we just discussed. Specifically, a marginal increase in wage causes a redistribution of consumption between the two occupations. To connect Eq. (2.16) to the risk premium, define the risk premium \( R(w) \) of entrepreneurs

\[ \mathbb{E} u (\pi (z, w)) = u (\mathbb{E} \pi (z, w) - R(w)). \tag{2.17} \]

The key observation is that the risk premium \( R(w) \) depends on the wage \( w \). After some manipulation detailed in Appendix B.2, we obtain the final form

\[ U'(\phi^m) = -\phi^m u'(w^m) w'(\phi^m) R'(w^m). \tag{2.18} \]
This expression decomposes the channel through which the increase in the number of entrepreneurs affects welfare. In particular, an increase in the number of entrepreneurs raises wage due to tighter labor market, \( w'(\phi) > 0 \), which then impacts the risk premium \( R'(w) \).

The final welfare impact is measured in utility term \( u'(w^m) > 0 \), and its total size depends on the number of entrepreneurs \( \phi^m > 0 \).

One thing that might be surprising in Eq. (2.18) is that the expansion of aggregate output \( Y'(\phi^m) \) does not show up. The reason is subtle. As detailed in Appendix B.2, although output does increase \( Y'(\phi^m) > 0 \), the agent who changes the occupation from worker to entrepreneur needs to be compensated by exactly the same amount as \( Y'(\phi^m) \) to be indifferent between the two occupations. Therefore, the marginal welfare gain of increase in the number of entrepreneurs does not come from output increase, but solely from the risk premium reduction.

A novel insight from Eq. (2.18) is that the elasticity of the risk premium \( R'(w^m) \) is a sufficient statistics for the existence of market failure and its direction. This is because \( R'(w^m) \) is the only term that has an ambiguous sign in Eq. (2.18). Therefore, the competitive market generates insufficient entrepreneurs \( U'(\phi^m) > 0 \) if and only if the risk premium can be reduced by increasing wage \( R'(w^m) < 0 \). Note that this is the formal description of the heuristic explanation that we provided, and reveals the channel through which the planner improves welfare.
Figure 2.1: The planner’s objective function and the distributions of the entrepreneurs’ profit $\pi(z, w)$.

2.3.2 Why are entrepreneurs insufficient $\phi^m < \phi^{op}$?

The previous section has shown that the elasticity of the risk premium is a sufficient statistics for the sign of market failure. A natural question is “what are the factors that determine the sign of the elasticity of the risk premium $R'(w^m)$?” To answer this question intuitively, suppose that the wage increases. As Fig. 2.1 shows, there are two forces that affect the risk premium $R(w)$.

On one hand, the distribution of the entrepreneurs’ consumption shifts to the left since the profit $\pi(z, w)$ is a decreasing function of the wage $w$. This shift moves the distribution to a more risk-averse region of the utility function, so it raises the risk premium $R(w)$.

On the other hand, the variance of the profit goes down. Such decline becomes salient if one takes the limit $w \to \infty$, in which case the profit goes to $\pi(z, w) \to 0$ for all productivity levels $z$. This effect reduces the risk premium $R(w)$. 

77
Therefore, the elasticity is determined by the relative importance of these two forces. Under the standard parameterization, i.e., CRRA utility and Cobb-Douglas production functions, the latter force always wins, i.e., $R'(w) < 0$ for all $w > 0$ including the market equilibrium wage $w^m$. However, this is not true in general. The next proposition states that there is a set of fundamentals $(u, f, G)$ that satisfy standard assumptions but generates an excess number of entrepreneurs.

**Proposition 2.4.** There exists a set of fundamentals $(u, f, G)$ such that $(u, f)$ is strictly increasing, strictly concave and satisfy Inada conditions and the resulting market equilibrium generates an excessive number of entrepreneurs compared to constrained efficient level $\phi^{cp} < \phi^m$.

**Proof.** See Appendix B.1.2. □

The proof provides an example in which a wage increase raises risk premium $R'(w^m) > 0$. Intuitively, such scenario is possible if agents suddenly become more risk averse when consumption is below some threshold. Under such utility function, the force that makes agents risk averse jumps up, but the force that reduces the variance of profit $\pi$ works smoothly. Given the intuition of Fig.2.1, such preference justifies the reduction of wage and therefore reduction of entrepreneurs $\phi^{cp} < \phi^m$. Mathematically, the proof in Appendix B.1.2 presents an example in which the utility function exhibits a discontinuity in the curvature. Thus, proposition 2.4 highlights the subtlety of the constrained inefficiency.
2.4 Concluding remarks

We have discussed the intuition of constrained inefficiency in the framework of competitive risky entrepreneurship. We have pointed out the difficulty of understanding the economic intuition of constrained inefficiency even in the simplest models in which the individual agents face one-dimensional optimization problem. We then present the competitive entrepreneurship model, whose market equilibrium is easy to understand, so that one can transparently see how and why the constrained planner’s intervention improves welfare. Future research should explore the constrained inefficiency in other types of models such as financial constraints.
Chapter 3

Intensive Margin of the Volcker Rule:  
Price Quality and Welfare

Sakai Ando and Misaki Matsumura
3.1 Introduction

The Volcker rule, named after Paul Volcker (Volcker [2010]), is one of the most important regulations after the Great Recession designed to prevent future financial crises. It limits the ability of banking entities as dealers to take risks on their own books by banning proprietary tradings and other activities described in the Final Rule.\(^1\) Although the necessity of restricting the excessive risk taking by banking entities is widely agreed, as discussed in Duffie [2012] and summarized in the Final Rule, the specific regulation targeted at the risk taking of dealers has invoked various concerns about its economic implications. Despite the importance of the policy and active public debates, there has been no theoretical analysis on the validity of these concerns. In particular, one of the immediate concerns is about the intensive margin (short-term impacts);\(^2\) the dealer regulation that ties the hands of price making institutions deteriorates the quality of price (informativeness and volatility) and therefore has a negative welfare impact.

This paper studies in a general equilibrium framework the robustness of this conventional wisdom about the intensive margin. Our main result is that the dealer regulation may lower price quality and dealer’s profitability as in the conventional wisdom, but at the same time if general equilibrium effect is taken into consideration, it may raise the welfare of other market participants against the conventional wisdom. In two extensions of the baseline model, we also argue that this novel insight on the welfare is robust, while the conventional

---

\(^1\)Prohibitions and Restrictions on Proprietary Trading and Certain Interests in, and Relationships with, Hedge Funds and Private Equity Funds, Office of the Federal Register, Vol.79, No. 21, National Archives and Records Administration.

\(^2\)The extensive margin, on the other hand, is mainly about the unknown consequences of the migration of dealers to unregulated areas. Given that the extensive margin has not materialized as Kelleher et al. [2016] argues, this paper focuses on the intensive margin.
wisdom about the deteriorating price quality is fragile to the introduction of dynamics and endogenous information acquisition.

Our welfare analysis deepens the debate on the dealer regulation by identifying the distribution of cost bearing. The typical discussion of dealer regulation focuses on the trade-off between the benefit from lower systemic risk and the cost due to efficiency loss caused by constraints imposed on dealers. This paper elaborates on the discussion of the cost by identifying dealers as the only group of market participants who bear the cost even though all the market participants face a less informative and more volatile price.

To demonstrate the seemingly counter-intuitive results rigorously, the baseline model formally describes the general equilibrium force that improves the welfare of non-dealers despite the price quality deterioration. The price quality deterioration that the baseline model captures is based on the same mechanism as the one widely discussed in public comments summarized in the Final Rule; if the dealer is not allowed to buffer temporary supply and demand imbalances, the price she quotes has to reflect those imbalances rather than economic fundamentals, and therefore becomes less informative and more volatile. Accordingly, one can show her expected profit decreases. The intuition behind the seemingly counter-intuitive welfare results is that, in a general equilibrium, somebody has to hold risky assets. If a dealer regulation restricts the dealer’s risk-taking ability, the risky assets held previously by the dealer have to be held by other market participants. The only way for a dealer to induce other market participants to hold risky assets is to quote an attractive price for them, resulting in a welfare redistribution from the dealer to other market participants. At the same time, for other market participants to accept more risks, this welfare redistribution has to be large enough to make them better off.
To critically evaluate the robustness of our mechanism, we relax the two simplifying assumptions in the baseline model in the two extensions. These exercises not only validate the robustness of our welfare results but also point out the fragility of the conventional wisdom about price quality. In the first extension, since the baseline model specifies the initial endowments of risky assets exogenously, which can drive the welfare results, we use the dynamic model to obtain the endogenous steady state inventories of risky assets. The analysis suggests that the welfare implications in the baseline model survive in the steady state, but the price volatility decreases against the conventional wisdom. In the second extension, we endogenize the precision of the signal about the economic fundamentals, which is the key ingredient in the baseline model to capture the price informativeness. The extension suggests that the welfare results of the baseline model are again robust, but the price informativeness might increase against conventional wisdom.

The Volcker rule and how we map it to a stylized model. In analyzing the effect of the Volcker rule, the obvious challenge is how to map the complicated actual regulation into an economic model. In particular, a large portion of the debates on the Volcker rule focuses on the difficulty of telling the proprietary trading from pure dealing. We argue that even if the regulators cannot cleanly detect proprietary tradings, the Volcker rule works as a deterrence device so that the impact of the Volcker rule, at least qualitatively, can be analyzed by considering the “effective” risk aversion of the dealer. In the following, we illustrate the institutional details of the Volcker rule and the justification for using the “effective” risk aversion of the dealer as the modeling device. For the complete description, see the Final Rule.

The Volcker rule is a section of the Dodd-Frank Wall Street Reform and Consumer
Protection Act. It bars banks from engaging in proprietary trading and having relationships with covered funds, with several exemptions including market making. To achieve this goal, the rule mandates that each bank under regulation runs a compliance program and requires big banks to report to regulators seven quantitative measures of trading activities, which are calculated every day for each individual trading desk.\(^3\) Enforcement tools include criminal and civil penalties,\(^4\) which are presumably used as a threat to incentivize banks to police themselves as part of their compliance program. Although the Volcker rule has not been applied against any cases yet, some argue that a good predictor of actual punishment procedure is the so-called London Whale trading mess, which eventually cost J.P. Morgan Chase $920 million in fines (Henning [2013]).

We map the implementation of the Volcker rule to the increase in the dealer’s effective risk aversion parameter. In section 3.2.3, we micro-found our comparative statics by showing how a severer threat of possible regulatory intervention can be mapped to a higher effective risk aversion parameter of the dealer.

One can also simply interpret that our choice of modeling describes an idealistic regulation; since the absolute goal of the Volcker rule is to reduce dealers’ excessive risk taking activities, the comparative statics in the dealer’s effective risk aversion is a direct thought experiment on its effect. Therefore, our analysis is robust to possible future changes in the actual implementation of the Volcker rule, and is applicable to potential similar dealer regulations outside of the U.S. that share the same goal.

In reality, however, due to the difficulty of detecting proprietary trading, the Volcker

\(^3\)See the Final Rule, III.D and III.E for overviews of metrics reporting and compliance program requirement.

\(^4\)See the Final Rule, IV.C.4. for other possible enforcement tools.
rule might not be able to perfectly achieve what it intends. In this case, one may be more interested in the effect of the Volcker rule when it can only achieve its goal to some extent, i.e., when the Volcker rule can eliminate some but not all risk takings on dealers’ own books. Our comparative statics in a continuum of the effective risk aversion levels can provide insights on all such intermediate cases. We provide further discussions in section 3.2.3.

**Literature.** To the best of our knowledge, this paper is the first to formally analyze the Volcker rule based on an internally consistent general equilibrium model, although the economic implications of the dealer regulation themselves have attracted interest after the financial crisis of 2007. The Final Rule summarizes public comments that reflect a wide range of opinions from both academia and industries. Duffie [2012] is one of the critical assessments of the Volcker rule based on various empirical and theoretical research. Trebbi and Xiao [2015] report that the Volcker rule has not produced structural deterioration in market liquidity. Kelleher et al. [2016] argues that the incumbent big dealers are still in their positions since they find legal loop holes and discourage new entrants. We complement these contributions by formalizing the mechanisms behind some of the concerns about the intensive margin, and pointing out a novel insight on welfare implications of the dealer regulation.

From the modeling point of view, our baseline model is an extension of Grossman and Stiglitz [1980] with a monopolistic and risk averse dealer who quotes the price, and our infinite horizon model is a descendant of Wang [1994]. Another paper that also extends Grossman and Stiglitz [1980] with a risk averse dealer is Liu and Wang [2016], which studies the bid-ask spreads under information asymmetry in a static model. As a modeling contribution, we show that a price making dealer can be cleanly embedded in both the static and dynamic frameworks, enabling us to isolate the effect of the dealer regulation from initial inventories.
of risky assets. See Vives [2010] for a coherent summary of the extensive literature on the dealership in general.

The negative correlation between informativeness and welfare of market participants in our model may remind some readers of the Hirshleifer effect, which describes that better information can make market participants worse off by destroying insurance opportunity. (Hirshleifer [1971]) However, as opposed to Hirshleifer effect, a change in informativeness is not the cause of a change in welfare in our model. Both price quality and welfare redistribution are the consequences of the dealer’s risk shifting. In this sense, our model points out a mechanism that is different from that of the Hirshleifer effect.

In this paper, we focus on the intensive margin, although the extensive margin is also a big portion of debates on the Volcker rule. For instance, Duffie [2012] discusses unpredictable consequences of dealer’s migration to unregulated sectors. A direct analysis of entry and exit in OTC markets can be found in Atkeson et al. [2015]. An interesting result of the three-type entry model of Atkeson et al. [2015] is that regulating dealer banks improves welfare, because in their model there is an excessive intermediation relative to socially optimal level. Although our analysis demonstrates that all other market participants than the dealer are better off, our mechanism is different from Atkeson et al. [2015]; we focus on welfare redistribution through price level, while this channel is muted in Atkeson et al. [2015], in which price is determined by Nash bargaining.
3.2 Baseline model

In this section, we describe the main intuition in a static model. We show as a result of dealer regulation, price quality deteriorates and dealer’s expected profit declines, but the welfare of other market participants improves.

3.2.1 Environment and definition

The baseline model contains two ingredients necessary to analyze the effects of dealer regulation on price quality and welfare: a price quoting dealer and a signal extraction problem. First, to formally think about the dealer’s pricing channel, it is necessary to introduce a dealer who quotes price optimally. As a consequence of introducing optimal pricing, the standard pricing mechanism based on demand and supply is extended to an optimal inventory management problem, in which the dealer’s optimal price pins down the trade-off between higher expected profit and riskier inventory, instead of clearing the excess demand.\footnote{This connection between a dealer’s inventory and her pricing has been emphasized in the literature (Amihud and Mendelson [1980], Ho and Stoll [1983], Treynor [1987], Grossman and Miller [1988], Hansch et al. [1998], and Liu and Wang [2016]) and became salient in the financial crisis of 2007 - 2009 where “the reduced dealer capacity resulted in dramatic downward distortions in corporate bond prices.” (Duffie [2012])}

Second, to discuss price informativeness, we introduce a signal extraction problem so that agents endogenously learn valuable information from equilibrium objects. In particular, we introduce heterogeneous traders such that some agents have private information about fundamentals. The dealer learns such private information through order flows, and quotes price optimally based on it. Since the price quoted is based on the extracted private information, the price itself is also informative about the fundamentals. Thus, the informativeness of equilibrium objects are determined endogenously.
Formally, there is one asset in the economy traded by three types of agents, insiders $I$ with mass $\lambda \in (0, 1)$, outsiders $O$ with mass $1 - \lambda$, and a dealer $D$. We call insiders and outsiders, combined, traders. Each player has a constant absolute risk aversion utility function (CARA), with parameters $\theta > 0$ for the traders and $\theta_D > 0$ for the dealer. The economy has three independent sources of uncertainty, which follow a common prior distribution

$$X = \begin{bmatrix} d \\ s \\ \epsilon \end{bmatrix} \sim N \begin{bmatrix} \bar{d} \\ \bar{s} \\ 0 \end{bmatrix}, \begin{bmatrix} \kappa_d^{-1} & 0 & 0 \\ 0 & \kappa_s^{-1} & 0 \\ 0 & 0 & \kappa_\epsilon^{-1} \end{bmatrix}.$$

(3.1)

$d$ denotes the return of the risky asset in the economy, which nobody can observe. $z := d + \epsilon$ is the signal on $d$ that only insiders can observe. $s$ is the dealer’s inventory of the asset, which only the dealer can observe. Risky inventory $s$ can be positive or negative, reflecting that the dealer can take either a long or short position. In the following, terminologies used to explain economic intuitions and interpretations assume the asset is valuable $\bar{d} > 1$ and inventory is positive $\bar{s} > 0$ on average, although symmetric arguments apply to other cases.

We work on the following equilibrium. Fix exogenous parameters $\{\theta, \theta_D, \bar{d}, \bar{s}, \kappa_d, \kappa_s, \kappa_\epsilon, \lambda\}$. Let $E_i$ be the conditional expectation operator conditional on the $\sigma$-algebra generated by the information set $\mathcal{F}_i$ of agent $i = I, O, D$.

**Definition 3.1.** A set of price and demand functions $\{p(z, s), x_I^B(z, p), x_O^B(p), x^H(z, p)\}$ constitutes an equilibrium if

1. (Traders) Demand functions $x_I^B(z, p)$ and $x_O^B(p)$ are best responses to the price quoted
by the dealer \( p(z,s) \).

\[
\begin{align*}
x_I^B(z,p) &= \arg \max_x E_I \left[ -e^{-\theta (d-p)x} \right], \quad \mathcal{F}_I = \{p(z,s), z\}, \\
x_O^B(p) &= \arg \max_x E_O \left[ -e^{-\theta (d-p)x} \right], \quad \mathcal{F}_O = \{p(z,s)\}.
\end{align*}
\tag{3.2}
\]

Aggregate demand satisfies \( x^B(z,p) = \lambda x_I^B(z,p) + (1 - \lambda) x_O^B(p) \).

2. (Dealer) Price quoted by the dealer \( p(z,s) \) is a best response to the aggregate demand \( x^B(z,p) \).

\[
\begin{align*}
p(z,s) &= \arg \max_{p} E_D \left[ -e^{-\theta_d \{s d + (p-d) x^B(z,p)\}} \right], \quad \mathcal{F}_D = \{s, x^B(z,p)\}.
\end{align*}
\tag{3.4}
\]

Note that it is straightforward to introduce a risk-free interest rate \( 1 + r \) and the initial wealth of the traders \( w_0 \). Without loss of generality due to CARA normal framework, \( r \) and \( w_0 \) can be set to be 0.

The interpretation of each agent’s problem is as follows. The traders are price takers and extract information about the return \( d \) from the price \( p(z,s) \). Given the information, the traders optimally choose the demand schedule \( p \mapsto x_i^B \), which sums up to the aggregate demand \( x^B(z,p) \). The superscript \( B \) represents best response. The dealer is a demand taker and extracts information about the return \( d \) from \( x^B(z,p) \). Based on the information, she quotes the price optimally by controlling the sum of the certain profit \( px^B(z,p) \) and the risky inventory \( \{s - x^B(z,p)\} d \). The equilibrium requires the quoted price to coincide with the price \( p(z,s) \) taken as given by the traders. Thus, the pair of price and demand \( \{p(z,s), x^B(z,p)\} \) that closes this loop constitutes a fixed point. In equilibrium, the dealer

89
can infer the signal $z$ from the demand schedule due to the affine structure stated in theorem 3.1, so that the outsiders are the only group of agents who do not observe the signal $z$.

The equilibrium objects can be used to define equilibrium trading volumes and welfare. The trading volumes of individual agents and the total trading volume in equilibrium can be obtained by substituting $p$ in demand functions with $p(z, s)$, i.e., $x_I(z, s) := x^B_I(z, p(z, s))$, $x_O(z, s) := x^B_O(p(z, s))$, and $x(z, s) := x^B(z, p(z, s))$. With these equilibrium objects, we can define the welfare of each agent. The welfare of the traders is measured by the ex-ante utility

$$u_i := E \left[ -e^{-\theta (d - p(z, s)) x_i(z, s)} \right], \ i = I, O. \quad (3.5)$$

For the dealer, since we conduct a comparative statics with respect to $\theta_D$ as explained in section 3.2.3, we adopt the expected profit as her welfare criterion to avoid mechanical welfare changes.

$$u_D := E \pi (d, z, s) = E \left[ sd + \{p(z, s) - d\} x(z, s) \right]. \quad (3.6)$$

The expected profit itself is of particular interest in the Final Rule, since some of the public comments show the concern that U.S. banks lose international competitiveness due to decreasing profitability.

### 3.2.2 Characterization of equilibrium

This section characterizes the affine structure of the equilibrium and defines the price informativeness.

The next theorem states that the equilibrium price and quantities are affine.
Theorem 3.1. There is a unique equilibrium such that the price function $p(z,s)$ is affine. In this equilibrium, the demand and the trading volume functions are also affine. The unique set of coefficients $\{\alpha_I, \beta_I, \gamma_I, \alpha_O, \beta_O, \alpha, \beta, \gamma, A, B, C\}$ of the equilibrium

\[ p(z,s) = A + B(z + Cs), \quad x^B(z,p) = \alpha + \beta p + \gamma z, \quad (3.7) \]

\[ x^B_I(z,p) = \alpha_I + \beta_I p + \gamma_I z, \quad x^B_O(p) = \alpha_O + \beta_O p \quad (3.8) \]

satisfies $BC\beta_I\gamma_I\beta_O\beta \neq 0$, $C < 0 < B$, and

\[ -\frac{\kappa_d + \kappa_e}{\theta} = \beta_I - \beta = \lambda\beta_I + (1 - \lambda) \beta_O < \beta_O < 0 < \gamma = \lambda\gamma_I < \gamma_I. \quad (3.9) \]

Proof. See Appendix C.1.  \hfill \Box

The signs of $(\beta_I, \beta_O, \beta)$ and $(\gamma_I, \gamma)$ reflect respectively the law of demand and the fact that a higher signal $z$ on the return $d$ pushes up demand of the asset. The difference in the price elasticity $|\beta_O| < |\beta_I|$ can be understood by noting that for outsiders, two forces are at work in the opposite directions; an increase in price not only reduces their demand by lowering the total return $d - p$, but also increases their demand by signaling the dealer’s information about the higher return $d$. Since the insiders observe $z$, they are only subject to the former force. As a result, the aggregate demand function $x^B(z,p)$, which is a convex combination of the traders’ demand functions, is more responsive to price than outsiders $x^B_O(p)$ but less than insiders $x^B_I(z,p)$. The positive sign of $B$ reflects the incentive for the

\(^6\text{Whether the affine price function } p(z,s) \text{ is unique in a larger set of functions, say, } C^1 \text{ or continuous functions, remains to be open. Papers about the uniqueness in models of Grossman and Stiglitz [1980] and Kyle [1985] include Boulatov et al. [2012], Palvolgyi and Venter [2015], and Breon-Drish [2015].} \)
dealer to raise her price when she faces a higher demand as a result of a better signal \( z \). The sign of \( C \) is negative since with a higher volume of inventory the risk averse dealer wants to cut the price to dispose of her inventory.

Eq. (3.7) in theorem 3.1 suggests a natural way to define price informativeness. For the outsiders, the term \( Cs \) is a noise that prevents them from inferring the signal \( z \) out of price \( p(z, s) \). When the variance of the noise term \( V(Cs) \) is large relative to that of the signal \( V(z) \), the variation of price function is dominated mainly by noise \( Cs \), so that the price is not informative about the valuable signal \( z \). In contrast, if \( C = 0 \), the price fully reveals \( z \).

Since the sign of \( C \) does not matter, we define the price informativeness as follows.

**Definition 3.2.** Consider the equilibrium characterized in theorem 3.1. The price informativeness \( Q \) is defined by

\[
Q := \frac{1}{|C|}. \tag{3.10}
\]

Now that we have equilibrium objects, we are ready to discuss the dealer regulation.

### 3.2.3 Mapping the dealer regulation to the model

This section explains how we map the Volcker rule into the stylized model and provides a formal micro-foundation as well as a verbal justification.

The way we map dealer regulation into the baseline model is to raise the dealer’s risk aversion parameter \( \theta_D \in (0, \infty) \). Such modeling choice reflects the interpretation that \( \theta_D \) represents the effective risk aversion rather than the deep preference parameter; due to the dealer regulation that bans the dealer from taking risks, the dealer behaves as if she becomes more risk averse. In particular, we analyze the comparative statics of (1) price
informativeness $Q$, (2) price volatility $V(p(z,s))$, and (3) welfare of agents $u_i$ for $i = I, O, D$, with respect to the dealer’s effective risk aversion $\theta_D \in (0, \infty)$.

As a micro-foundation, we justify our modeling choice by showing that controlling the risk aversion of the dealer is observationally equivalent to the dealer regulation with imperfect monitoring and pecuniary punishment. Suppose the dealer’s true risk aversion is $\theta > 0$ and when she makes her pricing decision she knows that she has to submit a report about her inventory to regulators, such as quantitative measures and annual certification by CEOs as specified in the Volcker rule. We assume that regulators can only get noisy information $m$ about the risky inventory

$$m(p) = s - x^B(z,p) + \xi$$

(3.11)

where the noise $\xi \sim N(0, \sigma^2_\xi)$ is independent of the uncertainty $X' = [d s e]$ and its accuracy $\sigma^2_\xi > 0$ can be chosen by regulators for some costs. Based on $m$, regulators impose a pecuniary punishment $F(m)$. Suppose regulators are reluctant to impose one when $|m|$ is close to 0 out of concern about false accusation, but are willing to impose a larger one when $|m| \to \infty$.

A simple parametrization of such punishment is $F(m) = \frac{\alpha}{2} m^2$ where $\alpha > 0$ represents how strictly regulators punish the dealer. In such setting, given the dealer regulation $(\sigma^2_\xi, \alpha)$, the dealer solves

$$\max_p -E_D e^{-\theta D} \left\{ (s-x^B(z,p))d + px^B(z,p) - \frac{\alpha}{2} (s-x^B(z,p) + \xi)^2 \right\}.$$  \hspace{1cm} (3.12)

Recall that the dealer’s problem (3.4) in the baseline model is

$$\max_p -E_D e^{-\theta D} \left\{ (s-x^B(z,p))d + px^B(z,p) \right\}.$$  \hspace{1cm} (3.13)
The following proposition states that the two problems are equivalent, and therefore justifies our parsimonious policy variable $\theta_D$.

**Proposition 3.1.** For each regulation $(\sigma_\xi^2, \alpha) \in \mathbb{R}^2_{++}$, there is an effective risk aversion $\theta_D \in (\theta, \infty)$ such that the prices quoted in problem (3.12) and (3.13) are identical. The converse is also true. For each effective risk aversion $\theta_D \in (\theta, \infty)$, there is a regulation $(\sigma_\xi^2, \alpha) \in \mathbb{R}^2_{++}$ such that the prices quoted in the two problems are identical.

**Proof.** See Appendix C.2. \hfill \square

An implication of proposition 3.1 is that we can always order the severity of regulations linearly. Such simplification not only makes the interpretations of the policy analyses simpler but also makes the results robust to the details of the implementation in the stylized model.

A verbal justification is to interpret that our analysis describes the ideal scenario of the Volcker rule, i.e., a scenario in which it can directly control the dealer’s risk-taking on her own books. In other words, we investigate the dealer regulation’s mechanics which are invariant to all possible implementations as long as they make the dealer effectively more risk averse. Thus, our analysis provides a benchmark result that is robust to possible future changes in the exact implementation of the Volcker rule, and is applicable to potential dealer regulations outside of the U.S. that share the same goal.

For such justification to be valid, it is desirable to show that the policy variable $\theta_D$ can fully span the situations of interest. Indeed, as the following proposition shows, by controlling $\theta_D$, we can capture both pure market making and pure proprietary trading.

**Proposition 3.2.** $\theta_D$ is a parameter that connects pure market making and pure proprietary trading. That is, for any $(z, s)$, $x(z, s) \to s$ as $\theta_D \to \infty$, and $x(z, s) \to 0$ as $\theta_D \to 0$. 

94
Proof. See Appendix C.3.

When the dealer becomes extremely risk averse \( \theta_D \to \infty \), all assets are held by the traders and therefore the dealer just makes market. This equilibrium is exactly the same as Grossman and Stiglitz [1980]. When \( \theta_D \to 0 \), the risk neutral dealer takes all the risks on her own book. The intermediate cases \( 0 < \theta_D < \infty \) correspond to the realistic situation where the Volcker rule deters proprietary trading only to some extent. Thus, by observing the entire range of \( \theta_D \), we can obtain policy-relevant insights even if the regulation is not as effective as is intended to be.

### 3.2.4 Results

The following theorem presents the analytical results. We also provide numerical results in Fig.3.1 to help illustrate the global behaviors of equilibrium objects of interest and their intuitions.

**Theorem 3.2.** Suppose the dealer becomes more risk averse. Price informativeness decreases. Price volatility eventually increases with sufficiently large inventory shocks. The welfare of the traders improves. The welfare of the dealer deteriorates. Formally,

1. The price informativeness \( Q \) is decreasing in \( \theta_D \) with \( \lim_{\theta_D \to \infty} Q = \frac{\lambda \kappa}{\theta} \) and \( \lim_{\theta_D \to 0} Q = \infty \).

2. Price volatility eventually increases, \( \lim_{\theta_D \to 0} V(p(z,s)) < \lim_{\theta_D \to \infty} V(p(z,s)) \), if \( \kappa_s \) is sufficiently small.

95
Figure 3.1: The four figures describe the impact of the increase in $\theta_D$ on equilibrium objects of interest. All the x axes are $\theta_D$. Parameter values are set to be $\kappa_d = \kappa_s = \kappa_e = \theta = \bar{s} = 1$, $\bar{d} = 1.1$ and $\lambda = 0.1$. Panel a shows that price informativeness deteriorates. Panel b shows that the price volatility eventually increases. Panel c shows that the welfare of traders improves, and the welfare difference widens. Panel d shows that the welfare of the dealer declines.

3. The welfare of the traders $\{u_I, u_O\}$ defined in (3.5) is higher when $\theta_D > 0$ than when $\theta_D \to 0$.

4. The welfare of the dealer $u_D$ defined in (3.6) is higher when $\theta_D \to 0$ than when $\theta_D \to \infty$.

Proof. See Appendix C.4.

The driving force behind theorem 3.2 and Fig.3.1 is the risk-shifting motive of the dealer. Note that the dealer’s objective function is equivalent to

$$
(s - x^B(z,p)) E[d|z] + px^B(z,p) - \frac{\theta_D}{2} (s - x^B(z,p))^2 V[d|z].
$$

(3.14)

The first two terms are the expected return of the inventory and the income from selling $x^B(z,p)$ units of the asset for the price $p$, and the third term governs the dis-utility from
holding risky assets. As she becomes more risk averse $\theta_D \to \infty$, the third term dominates
the dealer’s incentive so that she puts more weight on disposing her risky inventory, trying
to shift risky assets to other market participants.

With this in mind, the effect of the dealer regulation on price quality can be understood
in the same way as the conventional wisdom. When the dealer becomes more risk averse,
she is more interested in clearing her inventory, so that the information of $z$, which is useful
in raising expected profit but not in reducing the risk of her inventory, is reflected less in her
pricing decision. Therefore, the price she quotes becomes less informative about the return
d, and more about the inventory risk $s$, leading to lower price informativeness. Accordingly,
since the price has to move together with the inventory shock $s$ to reduce the inventory risks,
if the volatility of inventory $Vs = \kappa_s^{-1}$ is high enough, price volatility eventually increases.

One force that could confound this intuition is the rise in the traders’ price sensitivity $|\beta|$, which contributes to the non-monotonicity of panel b. of Fig.3.1. Recall $|\beta_O| < |\beta_I|$ since the
outsiders infer higher signal $z$ from higher price $p$. As the dealer’s risk aversion $\theta_D$ increases,
price informativeness decreases, so that outsiders no longer infer signal $z$ from the price
function $p(z, s)$. As a result, the outsiders’ price sensitivity $|\beta_O| = |\partial x^B_O (z, p) / \partial p|$ increases,
which then raises the price sensitivity of aggregate demand $|\beta| = |\partial x^B (z, p) / \partial p|$. This
higher responsiveness $|\beta| = |\partial (s - x^B (z, p)) / \partial p|$ makes it easier for the dealer to adjust
her inventory, so that the dealer may not have to change price as much to control inventory.
Although this force creates the non-monotonicity, it is of second order importance in the
limit.

What is new relative to the conventional wisdom is the result on the traders’ welfare.
This welfare result can also be explained by the shift of the dealer’s incentive. Since the
dealer cannot absorb shocks using her own inventory due to additional risk aversion, some
of the risks have to be shifted to the traders. The only instrument the dealer has is the
price. Therefore, in order for her to shift risks to the traders, she has to compensate them
for their riskier positions by quoting a more attractive price than ever before. For the traders
to accept riskier positions, this benefit has to be large enough to make them better off. As a
result, welfare is redistributed from the dealer to the traders to the extent that the traders
are better off. Accordingly, the dealer’s profitability deteriorates in exchange of less riskier
inventory.

One can also see in Fig. 3.1.c that the traders’ utility difference $u_I - u_O$ increases. This
is a corollary of the deteriorating price informativeness. Since price is less informative as
$\theta_D \to \infty$, the informational advantage of the insiders over the outsiders increases, which
is then reflected in the welfare difference. In other words, $u_I - u_O$ measures the value of
information, and it increases as the dealer becomes more risk averse.

3.2.5 Discussion

We discuss two potential concerns about the assumptions embedded in the baseline model:
(1) the exogenous initial endowments and (2) the exogenous information acquisition. These
assumptions are relaxed in the following extensions so that the robustness or the fragility of
the results in the baseline model are examined.

First, in the baseline model, the traders are not endowed with any risky asset for parsimony.
However, such simplification implies that the immediacy of trading comes from the
dealer. If the traders are endowed with a large amount of risky assets, the immediacy of
trading comes from the traders, in which case the dealer regulation might make tradings harder and deteriorate the welfare of the traders. To address such concern, in section 3.3, we endogenize the risky asset holdings before transactions by deriving the steady state in a dynamic model. In particular, we extend Wang [1994] to build a dynamic model with steady state risky asset holdings on both the dealer’s and the traders’ sides, and conduct an analogous policy analysis of the dealer regulation.

Second, we like to address the exogenous information acquisition by the insiders. In the baseline model, the precision of the signal $\kappa_{\epsilon}$ is exogenous. However, when the risk attitude of the dealer changes, the insiders, knowing there will be less information in equilibrium, might have incentives to strengthen their information acquisition activity. To see if the results in the baseline model are robust to the introduction of the endogenous information acquisition, in section 3.4, we allow $\kappa_{\epsilon}$ to depend on $\theta_D$, and conduct the same policy analysis as in the baseline model.

### 3.3 Dynamic model

This section argues that the welfare results of the baseline model survive even when the exogenous initial endowments are endogenized as the steady state of a dynamic model. We also show that the increase in price volatility, which is a part of the conventional wisdom derived in the baseline model, does not survive.

As discussed in section 3.2.5, the exogenous endowments in the baseline model has to be endogenous to deal with the concerns about the immediacy of tradings. For this purpose, we extend the baseline model along Wang [1994] by adding a price-making dealer.
### 3.3.1 Environment and definition

There are one price-making dealer $D$ and a continuum of identical price-taking informed traders $I = [0,1]$.\(^7\) Their preference parameters are the rate of absolute risk aversion $\theta_i$, $i = D,I$ and the common discount factor $\beta \in (0,1)$. They trade a single risky asset $x_t$ with risky dividend $d_{t+1}$ and a risk-free bond $y_t$ with gross interest rate $R$. The total supply of the risky asset is $\bar{x}$, and the supply of the risk-free asset is inelastic so that $R$ is exogenously given. Each agent $i = D,I$ has a private investment opportunity $s^i_t$, generating a risky dividend $d^i_{t+1}$. At each period $t$, each agent divides revenue $p_t x^i_{t-1} + d_t x^i_{t-1} + R y^i_{t-1} + s^i_{t-1} d^i_t$ into profit $\pi^i_t$, investment in the risk-free bond $y_t$ and investment in the risky asset $p_t x^i_t$. Hence, the flow budget for each agent $i = D,I$ can be written as

$$\pi^i_t + y^i_t + p_t x^i_t = p_t x^i_{t-1} + m^i_t, \quad m^i_t := d_t x^i_{t-1} + R y^i_{t-1} + s^i_{t-1} d^i_t$$  \hspace{1cm} (3.15)$$

where $m^i_t$ is the amount of money at the beginning of period $t$ that is independent of the period $t$ price.

To define the equilibrium, we need to specify the information set of each agent. As in the baseline model, the informed traders receive a noisy signal about the dividend $z_t = d_{t+1} + \epsilon_{t+1}$. The information set for the informed traders $\mathcal{F}^I_t = \{ p_{\tau-1}, x^I_{\tau-1}, d_{\tau}, d^I_{\tau}, s^I_{\tau}, z_{\tau} \}_{\tau \leq t}$ contains all the past prices, asset holdings, dividends, private investment opportunities and signals. Given the prices $(p_t,R)$ and the information $\mathcal{F}^I_t$, the informed traders submit the demand schedule $x^I_t = x^I_t (z_t, s^I_t, p_t)$ where the notation emphasizes that the demand schedule is a noisy

---

\(^7\)Since both the dealer and the traders have private shocks, we do not need to introduce uninformed traders to prevent full information revelation.
signal about \( z_t \). The information set for the dealer \( \mathcal{F}_t^D = \{ p_{r-1}, x_{r-1}^D, d_r, d_r^D, s_r, x_r^I (z_r, s_r^I, p_r) \}_{\tau \leq t} \) contains the demand schedule \( x_t^I (z_t, s_t^I, p_t) \) so that the dealer extracts information about the signal \( z_t \). The dealer quotes price optimally knowing that her inventory is determined by 
\[
x_t^D (p_t) := \bar{x} - x_t^I (z_t, s_t^I, p_t).
\]

We assume the exogenous uncertainty \( X_t = [d_t, d_t^D, d_t^I, s_t, s_t^D, \epsilon_t]' \) is \( i.i.d. \) over time \( t \) and is joint normal

\[
X_t \sim N \left( \begin{bmatrix} \bar{d} \\ \bar{d}^D \\ \bar{d}^I \\ 0_{3\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_{3\times 3} & 0 & 0 & 0 \\ 0 & \sigma_{sI}^2 & 0 & 0 \\ 0 & 0 & \sigma_{sD}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\epsilon}^2 \end{bmatrix} \right). \tag{3.16}
\]

The correlation among \((d_t, d_t^D, d_t^I)\), denoted by the \( 3 \times 3 \) matrix \( \Sigma_{3\times 3} \), generates reasons to trade, and the \( i.i.d. \) assumption suffices to yield AR(1) equilibrium asset holdings as stated in proposition 3.3.

The equilibrium is defined as follows. Let \( E_t^i \) be the expectation operator conditional on the \( \sigma \)-algebra generated by \( \mathcal{F}_t^i, i = D, I \).

**Definition 3.3.** The sequence of asset holdings and prices \((x_t^D, p_t)_t\) is an equilibrium if the following conditions are satisfied.

1. At each \( t \), given the information \( \mathcal{F}_t^I \) and prices \((p_t, R)\), the traders’ asset holding 
\[
x_t^I (z_t, s_t^I, p_t)
\]

solves

\[
J_t^I = \max_{\pi, x, y} -E_t^I \sum_{u=0}^{\infty} \beta^u e^{-\theta \tau_{t+u}} \text{s.t.} \begin{cases} m_t^I = d_t x_{t-1}^I + R y_{t-1} + s_{t-1}^I d_t^I \\
\pi_t + y_t + p_t x_t = p_t x_{t-1}^I + m_t^I \end{cases}.
\]

101
2. At each $t$, given the information $\mathcal{F}_t^D$ and the price $R$, the dealer’s price $p_t$ solves

$$J_t^D = \max_{\pi, p, y} -E_t^D \sum_{u=0}^{\infty} \beta^u e^{-\theta_D u} \pi_t + u \text{ s.t. } \begin{cases} m_t^D = d_t x_{t-1}^D + R y_{t-1} + s_{t-1}^D p_t^D \\ \pi_t + y_t + px_t^D = px_{t-1}^D + m_t^D \\ x_t^D = \bar{x} - x_t^I (z_t, s_t^I, p) \end{cases}.$$ (3.18)

Note that this equilibrium shares the same spirit as the baseline model; the dealer is a demand taker and the traders are price takers. In this sense, the price and the demand are best responses to each other.

The main difference is that since the problem is dynamic, the risky asset purchase of the last period is directly tied to the current period risky inventory. Therefore, both agents can control their inventory shocks as opposed to the baseline model.

3.3.2 Characterization of equilibrium

This section characterizes the affine equilibrium and shows that the dynamic model has a well-defined steady state distribution of the risky asset holdings.

Fix the following sixteen exogenous parameters

$$\{\beta, R, \theta_I, \theta_D, \bar{x}, \bar{d}, \bar{d}^I, \bar{d}^D, \sigma_d, \sigma_e, \sigma_{ddI}, \sigma_{ddD}, \sigma_{dI}, \sigma_{dD}, \sigma_{sI}, \sigma_{sD} \}$$ (3.19)

where $(\bar{d}, \bar{d})$ are the means of $(d_t, d_t^I), (\sigma_d, \sigma_e, \sigma_{ddI}, \sigma_{ddD}, \sigma_{dI}, \sigma_{dD}, \sigma_{sI}, \sigma_{sD})$ are the variances of $(d_t, \epsilon_t, d_t^I, d_t^D, s_t^I, s_t^D)$, and $(\sigma_{ddI}, \sigma_{ddD})$ are the covariances of $(d_t, d_t^I)$ and $(d_t, d_t^D).$\(^8\)

\(^8\)We do not have to specify the covariance of $(d_t^I, d_t^D)$ since it does not affect individual optimization, and therefore equilibrium objects.
Proposition 3.3. There is an affine equilibrium \((x_t^D, p_t)\), i.e., for some constants of price \(A_0, A_x, B, C = [C_l, C_D]'\) and constants of risky asset holdings \((\rho_0, \rho_1)\), the equilibrium has the form of

\[
p_t = A_0 + A_x x_{t-1}^D + B (z_t + C' s_t), \quad s_t = [s_t^l, s_t^D]', \tag{3.20}
\]

\[
x_t^D = \rho_0 + \rho_1 x_{t-1}^D + \epsilon_t^D , \quad \epsilon_t^D \sim N \left(0, V\epsilon_t^D\right), \tag{3.21}
\]

where \(\epsilon_t^D\) is a function of current shocks \((z_t, s_t)\).

Proof. See Appendix C.5. \qed

This proposition satisfies our motivation to introduce dynamics; if \(|\rho_1| < 1\), we can obtain the average asset holding

\[
E x^D := \frac{\rho_0}{1 - \rho_1}. \tag{3.22}
\]

Another observation is that the equilibrium has the state-space representation where Eq. (3.21) is the state equation and Eq. (3.20) is the observation. The reason for which the equilibrium objects are persistent despite the \(i.i.d.\) shock assumption is the inventory management. The transaction of the last period on the risky asset affects the amount of risks to begin with in the current period, which then affects the transaction in the current period.

With the affine equilibrium structure, we can conduct an analogous policy analysis as the baseline model.

3.3.3 Policy analysis

As in the baseline model, we focus on the comparative statics of equilibrium objects with respect to the effective risk aversion \(\theta_D \in (0, \infty)\). The interpretation of the thought expe-
riment is as follows. Suppose the economy is in the steady state. If suddenly the dealer regulation is introduced, what will happen to the price quality and welfare? In particular, we note that our welfare analysis takes into account the transitional dynamics, not just the comparison of the steady states before and after the policy intervention.

Let us formalize the equilibrium objects of interest. The price informativeness is defined as $|C_D|^{-1}$ which reflects how much the information that the dealer extracts from the order flow goes into the equilibrium price. We report the conditional and unconditional variances $V(p_t|x_{t-1}^D)$ and $V(p_t)$ as price volatility. The welfare of the traders is defined as $E[J_I^0|x_{t-1}^D,m_I^0]$ where the asset holding is the steady state value derived in Eq. (3.22) when $\theta_D = \theta_I$, denoted by $x_{t-1}^D = E x^D (\theta_I = \theta_D)$, and the initial money is set to be $m_{I_0} = 0$ without loss of generality. For the welfare of the dealer, we see the path of expected profit $t \mapsto E[\pi_t^D|x_{t-1}^D,m_{D_0}^D]$, again with $x_{t-1}^D = E x^D (\theta_I = \theta_D)$ and $m_{D_0}^D = 0$. Fig.3.2 shows a numerical example where the parameter values are specified in the caption.

What is different from the baseline model is the declining price volatility, which increases according to the conventional wisdom. The logic of decreasing price volatility can be understood from observing panel f. As the dealer becomes more risk averse, the dealer reduces the risky asset holding. Since the absolute amount of risk in her inventory decreases, she does not have to fluctuate price as much as in the baseline model where the inventory risks from the endowments are exogenously fixed. In other words, since in the dynamic model the inventory risks can be endogenously chosen to be small by reducing the risky asset holding, there is less need to make price fluctuate.

All other results in the baseline model including the welfare implications survive in this dynamic setting. Price informativeness deteriorates, the welfare of the traders improve and
Figure 3.2: The parameters are $\beta = R^{-1} = .9$, $\theta_I = 1$, $\bar{d} = 2$, $\bar{d}^I = \bar{d}^D = 1$, $\sigma_d^2 = \sigma_t^2 = \sigma_d^2 = \sigma_d^D = \sigma_d^2D = \sigma_d^2I = \sigma_d^2 = \sigma_d^2 = .5$ and $\bar{x} = 200$ so that the steady state asset holdings are positive for both the dealer and the traders. All the $x$ axes except for panel d are $\theta_D$. The $x$ axis for panel d is time $t$. Panel a shows that the price informativeness deteriorates as a result of dealer regulation from the steady state. Panel b and c show that price volatility decreases as opposed to Fig. 3.1. Panel c shows that the traders’ welfare improves. Panel d shows that the expected profit of the dealer declines at each time horizon. Panel f shows that as the dealer becomes more risk averse, the steady state asset holding for the dealer decreases.

the welfare of the dealer decreases. In particular, the expected profit for the dealer decreases not just in the sense of the discounted sum, but also at each future period.

In summary, we observe that the welfare results in the baseline model are robust, but the increase in price volatility is flipped by introducing dynamics, suggesting the fragility of the conventional wisdom.
3.4 Endogenous information acquisition

The section shows that the welfare results of the baseline model survive even when information acquisition is endogenized. We also show that the decreasing price informativeness, derived in the baseline model as a description of the conventional wisdom, does not survive.

3.4.1 Definition of equilibrium

To endogenize the information acquisition, we construct a simple two-period model, in which the insiders choose $\kappa_\epsilon$ in the first period, and all players play the baseline model in the second period. Hence, we use the same notation as the baseline model. Let $u_I(\kappa_\epsilon)$ be the equilibrium ex-ante utility of the insiders in the baseline model, viewed as a function of the precision of the signal $\kappa_\epsilon$.

Definition 3.4. Fix exogenous parameters $\{\theta, \theta_D, \bar{d}, \bar{s}, \kappa_d, \kappa_s, \lambda\}$. A set of signal precision, price and demand functions $\{\kappa^*_\epsilon, p(z,s), x^B_I(z,p), x^H_O(p), x^B(z,p)\}$ is a subgame perfect equilibrium if

1. $\kappa^*_\epsilon$ maximizes $u_I(\kappa_\epsilon)$.

2. Given $\{\theta, \theta_D, \bar{d}, \bar{s}, \kappa_d, \kappa_s, \kappa^*_\epsilon, \lambda\}, \{p(z,s), x^B_I(z,p), x^H_O(p), x^B(z,p)\}$ constitutes an equilibrium in the baseline model.

This definition describes the best case scenario for the insiders in the sense that the insiders are allowed to take the social planner’s point of view, understanding all the general equilibrium effects in the second period when they optimize the signal precision $\kappa_\epsilon$ in the first period.
A natural concern about this equilibrium concept is whether the equilibrium precision \( \kappa^*_\epsilon \geq 0 \) is an interior solution or not. In the current setting, there is a natural trade-off that keeps \( \kappa^*_\epsilon \) interior. As the precision of the signal \( \kappa_\epsilon \) increases, the insiders are more certain about the distribution of the return \( d \). This partial equilibrium effect has a positive influence on the ex-ante utility of the risk averse insiders. In contrast, an increase in \( \kappa_\epsilon \) also informs the dealer of a more precise signal \( z \) in equilibrium, which reduces the dealer’s inventory risk and therefore allows her to exert her monopolistic power more confidently. This reduces the insiders’ welfare. Put differently, a better signal decreases the conditional variance \( V[d|z] \) in Eq.(3.14), affecting the dealer’s pricing decision in the same way as a decrease in \( \theta_D \). The resulting higher price then adversely affects the insiders’ welfare. Given this trade-off, we can numerically solve the interior solution \( \kappa^*_\epsilon > 0 \) for each \( \theta_D \) without introducing specific information cost such as rational inattention a la Sims [2010] or hiring analysts.

Another observation is that this definition abstracts from strategic interactions among the insiders. One can micro-found the first stage optimization as an efficient Nash equilibrium. Suppose that the highest precision chosen among all the insiders becomes the precision of the signal observed in the second period. When other insiders choose the precision \( \kappa^*_\epsilon \), choosing \( \kappa^*_\epsilon \) is optimal. One can also introduce an infinitesimal cost of information acquisition to this micro-foundation. In this case, the unique pure strategy Nash equilibrium is that one agent chooses \( \kappa^*_\epsilon \) and all others free-ride. Since the motivation behind the exercise is not to micro-found information acquisition, but to see if the baseline results are robust to information acquisition, we simplify the strategic interactions by focusing on the efficient information acquisition that maximizes the insiders’ welfare.
3.4.2 Results and discussion

The same four equilibrium objects as the baseline model are of interest, i.e., price informativeness $Q$, price volatility $V(p(z,s))$, the welfare of the traders $\{u_I, u_O\}$, and the welfare of the dealer $u_D$. The only difference from the baseline model is that $\kappa_e$ is a function of $\theta_D$ since $\kappa_e$ is a choice variable. Fig. 3.3 summarizes the numerical findings.

The most salient contrast with the baseline model is the improving price informativeness, which shows that the insiders have a stronger incentive to pump more information into the market as the dealer becomes more risk averse. The intuition behind this result is that as the dealer becomes more obsessed with her inventory, the insiders can raise the precision of the signal to enjoy the partial equilibrium effect without being exploited by the dealer’s monopolistic power. The fact that the price informativeness increases once the insiders’
information acquisition is endogenized suggests that the effect of dealer regulation on price informativeness could potentially go either way, as opposed to what the conventional wisdom suggests.

Other panels follow the same patterns as in the baseline model, including the welfare redistribution from the dealer to the traders. In particular, the policy analysis indicates the robustness of the welfare results in the baseline model.

3.5 Final Remarks

We have analyzed the effects of dealer regulation on the properties of price and the resulting welfare consequences. The baseline model shows that the price quality deterioration can coexist with the welfare improvement of other market participants than the dealer. The two extensions then demonstrates the robustness of the novel welfare implications as well as the fragility of the conventional wisdom on the price quality deterioration. We are going to conclude the paper by describing other important aspects of dealer regulation that this paper does not address.

In this paper, we have limited our scope to the intensive margin of dealer regulation. However, considering extensive margin is also imperative for a comprehensive assessment of the Volcker rule. Although Kellcher et al. [2016] reports that the migration has not happened due to the efforts by the incumbent dealers to discourage entrance, as emphasized in Duffie [2012], “a potential migration of market making to the outside of the regulated bank sector might have unpredictable and potentially important adverse consequences for financial stability.” See Whitehead [2011] for more discussion.
Although we have described welfare implication for the intensive margin in a stylized model, a quantitative assessment requires a comprehensive cost-benefit analysis. Challenges for a comprehensive quantitative analysis of the Volcker rule includes an identification strategy to filter out confounding factors, constructing a counter-factual of a rare event in which the Volcker rule mitigates a financial crisis, quantifying the benefit of reducing the dealers’ conflicts of interests, evaluating the cost of losing the competitiveness of U.S. banking entities as dealers, and other important issues listed in the Final Rule. These are the points worth more investigation as well as the important caveats in understanding the results of this paper.


Julio Davila, Jay H. Hong, Per Krusell, and Jose-Victor Rios-Rull. Constrained Ef-


Emmanuel Farhi, Mikhail Golosov, and Aleh Tsyvinski. A Theory of Liquidity and Regula-


Sanford J. Grossman and Joseph E. Stiglitz. On the Impossibility of Informationally Efficient


Cesaire A. Meh. Business risk, credit constraints, and corporate taxation. Journal of Econo-


Alexis Akira Toda. Asset prices and efficiency in a Krebs economy. Review of Economic


Appendices
Appendix A

Appendix to Chapter 1

A.1 Proofs of propositions

I use the following lemma.

**Lemma A.1.** *(Jensen’s inequality)* Suppose $X > 0$ is a random variable with positive variance $VX > 0$. Then, for any $\gamma > 0$,

$$(E X^{1-\gamma})^{\frac{1}{1-\gamma}} < E X.$$

**Proof.** Since CRRA utility function is strictly concave and increasing when $\gamma > 0$, Jensen’s inequality and taking inverse function imply

$$E u (X) < u (E X) \iff u^{-1} (E u (X)) < E X \iff (E X^{1-\gamma})^{\frac{1}{1-\gamma}} < E X.$$
A.1.1 Proof of proposition 1.2

The planner’s solution $\phi^P$ maximizes the output.

$$Y(\phi, n) = \max_{n(z) \geq 0, \phi \in [0,1]} \phi E[z f(n(z)) | s \geq G_s^{-1}(1 - \phi)]$$

s.t. $\phi + \phi E[n(z) | s \geq G_s^{-1}(1 - \phi)] = 1.$

Reformulate the problem using $\bar{s}$ gives

$$Y(\bar{s}, n) = \max_{\bar{s}, n(z) \geq 0} \int zf(n(z)) 1_{s \geq \bar{s}} dG \quad s.t. \int n(z) 1_{s \geq \bar{s}} dG = G_s(\bar{s}).$$

For each fixed $\bar{s}$, the sub-problem of choosing $n(z)$ is

$$\max_{n(z) \geq 0} \int zf(n(z)) g(z|s \geq \bar{s}) dz \quad s.t. \int n(z) g(z|s \geq \bar{s}) dz = \frac{G_s(\bar{s})}{1 - G_s(\bar{s})}.$$ 

The objective function is a weighted sum of the strictly concave function of $\{n(z)\}_z$ and the constraint is a linear function of $\{n(z)\}_z$. Hence, the unique solution is characterized by the first order conditions.

Furthermore, let $s_{\min}, s_{\max} \in [-\infty, \infty]$ be the smallest and largest values that the signal can take, so that $1_{s \geq s_{\min}} = 1$ and $1_{s \geq s_{\max}} = 0$ for all realizations. When $\bar{s} = s_{\min}$, the value is $Y(s_{\min}, n) = 0$ since the only feasible employment schedule is $n(z) = 0$ for all $z$. When $\bar{s} = s_{\max}$, $Y(s_{\max}, n) = 0$ since the objective function is 0. Since 0 output is the lower bound $Y \geq 0$, the solution $\bar{s}$ is not at the boundaries. Therefore, the solution exists and is unique if there is a unique set of $(\bar{s}, n)$ that solves the first order conditions.
The first order condition can be obtained by taking the derivative of the Lagrangian.

\[
\begin{align*}
\begin{cases}
  n(z) = (f')^{-1}\left(\frac{\lambda}{z}\right), & \lambda > 0. \\
  \mathbb{E}[zf(n(z)) | s = \bar{s}] = \lambda (1 + \mathbb{E}[n(z) | s = \bar{s}]) \\
  \mathbb{E}[n(z) 1_{s \geq \bar{s}}] = G_s(\bar{s})
\end{cases}
\end{align*}
\]

Combining the conditions gives

\[
\begin{align*}
\begin{cases}
  F_1(\bar{s}, \lambda) = \mathbb{E}[zf\left((f')^{-1}\left(\frac{\lambda}{z}\right)\right) - \lambda (f')^{-1}\left(\frac{\lambda}{z}\right) | s = \bar{s}] - \lambda = 0. \\
  F_2(\bar{s}, \lambda) = \mathbb{E}[(f')^{-1}\left(\frac{\lambda}{z}\right) 1_{s \geq \bar{s}}] - G_s(\bar{s}) = 0
\end{cases}
\end{align*}
\]

Note that the integrand of \(F_1\) is identical to the value function of the profit maximization problem with wage \(\lambda\)

\[
zf\left((f')^{-1}\left(\frac{\lambda}{z}\right)\right) - \lambda (f')^{-1}\left(\frac{\lambda}{z}\right) = \max_{n \geq 0} zf(n) - \lambda n.
\]

Thus, \(F_1\) gives a strictly positive relationship between \(\bar{s}\) and \(\lambda\). In contrast, \(F_2\) gives a strictly negative relationship. Hence, there is a unique solution \((\bar{s}, n)\) to the output maximization problem, and unique solution \(\phi^P\) to the original problem.
A.1.2 Proof of proposition 1.3

For the first statement, when there are no risks, the laissez-faire economy is characterized by \((\bar{z}^{LF}, w^{LF}, \pi^{LF}, n^{LF})\) that satisfy

\[
\begin{align*}
\pi \left( \bar{z}^{LF}, w^{LF} \right) &= w^{LF} \\
zf' \left( n^{LF} \left( z, w^{LF} \right) \right) &= w^{LF} \\
\pi \left( z, w^{LF} \right) &= zf \left( n \left( z, w^{LF} \right) \right) - w^{LF} n \left( z, w^{LF} \right) \\
\mathbb{E} \left[ n \left( z, w^{LF} \right) 1_{z \geq \bar{z}^{LF}} \right] &= G_z \left( \bar{z}^{LF} \right)
\end{align*}
\]

Choosing the Pareto weight using these variables.

\[
\lambda \left( z \right) = \begin{cases} 
\frac{1}{u'(w^{LF})} & z \leq \bar{z} \\
\frac{1}{u'(\pi(z, w^{LF}))} & z > \bar{z} 
\end{cases}
\]

I show that \((\bar{z}, w, \pi, n)\) is the solution to the planner’s problem. Without loss of generality, \(c_i(\omega) = c(z_i)\). The planner’s solution can be written as

\[
\max_{c,n,Y,\bar{z}} \int \lambda \left( z \right) u \left( c \left( z \right) \right) dG_z \text{ s.t. } \begin{cases} 
\int c \left( z \right) dG_z = Y \\
Y = \mathbb{E} \left[ zf \left( n \left( z \right) \right) 1_{z \geq \bar{z}} \right] \\
\mathbb{E} \left[ n \left( z \right) 1_{z \geq \bar{z}} \right] = G_z \left( \bar{z} \right)
\end{cases}
\]
The Lagrangian is
\[ L = \int \lambda(z) u(c(z)) \, dG_z + \mu \{ \mathbb{E} [zf(n(z)) \mathbf{1}_{z \geq \bar{z}}] - \mathbb{E} c(z) + \nu (G_z(\bar{z}) - \mathbb{E} [n(z) \mathbf{1}_{z \geq \bar{z}}]) \} \]

The FOCs are
\[
\begin{align*}
L_c &= 0 \implies u'(c(z)) = \frac{\mu}{\lambda(z)} \\
L_n &= 0 \implies zf'(n(z)) = \nu \\
L_{\bar{z}} &= 0 \implies \bar{z} f(n(\bar{z})) - vn(\bar{z}) = \nu
\end{align*}
\]

\((\mu, \bar{z}, v, n) = (1, \bar{z}^{LF}, w, n^{LF})\) and \(c(z) = \pi^{LF}(z, w^{LF}) \mathbf{1}_{z \geq \bar{z}^{LF}} + w \mathbf{1}_{z < \bar{z}^{LF}}\) satisfies the equilibrium conditions.

For the second statement it suffices to show that the laissez-faire economy does not maximize output. With risks, the laissez-faire economy is characterized by

\[
\mathbb{E} \left[ u(zf(h(w/z)) - wh(w/z)) \mathbf{1}_{s = \bar{s}} \right] = u(w) \\
\mathbb{E} \left[ h(w/z) \mathbf{1}_{s \geq \bar{s}} \right] = G_s(\bar{s})
\]

where \(h(x) := (f')^{-1}(x)\). The FOCs of the planner’s problem

\[
\max_{\bar{s}, n(z)} \int zf(n(z)) \mathbf{1}_{s \geq \bar{s}} dG \quad s.t. \quad \int n(z) \mathbf{1}_{s \geq \bar{s}} dG = G_s(\bar{s}),
\]

131
using the Lagrange multiplier $\lambda$, is

$$
\mathbb{E} \left[ zf \left( h \left( \frac{\lambda}{\bar{z}} \right) \right) - \lambda h \left( \frac{\lambda}{\bar{z}} \right) | s = \bar{s} \right] = \lambda
$$

$$
\mathbb{E} \left[ h \left( \frac{\lambda}{\bar{z}} \right) 1_{s \geq \bar{s}} \right] = G_s (\bar{s}).
$$

Note that the two systems are different by $u$. Therefore, the laissez-faire economy can Pareto improve by producing more and redistribute them.

The two systems also imply $w < \lambda$. To see this, note that Jensen’s inequality implies

$$
\lambda = \mathbb{E} \left[ zf \left( h \left( \frac{\lambda}{\bar{z}} \right) \right) - \lambda h \left( \frac{\lambda}{\bar{z}} \right) | s = \bar{s} \right] > u^{-1} \left( \mathbb{E} \left[ u \left( zf \left( h \left( \frac{\lambda}{\bar{z}} \right) \right) - \lambda h \left( \frac{\lambda}{\bar{z}} \right) | s = \bar{s} \right) \right] \right).
$$

In addition,

$$
h' (\lambda) < 0, \quad \frac{\partial}{\partial \lambda} \left\{ zf \left( h \left( \frac{\lambda}{\bar{z}} \right) \right) - \lambda h \left( \frac{\lambda}{\bar{z}} \right) \right\} = -h \left( \frac{\lambda}{\bar{z}} \right) < 0.
$$

so from the market clearing condition, $\lambda$ and $\bar{s}$ have to be negatively correlated. Hence, $w < \lambda$ and $\bar{s}^P < \bar{s}^{LF}$, which then implies $\phi^{LF} < \phi^P$. 

132
A.1.3 Proof of proposition 1.4

First, I reformulate the optimal taxation problem into a mechanism design problem of choosing a set of menu.

\[
\begin{align*}
\max_{w,C(z),Y(z)} w & \text{ s.t.} \\
\mathbb{E} u (zf (C(z)) - wC(z) - Y(z)) &= u(w) \\
U(C(z), Y(z), z) &\geq U(C(z'), Y(z'), z) \quad \forall z, z' \\
\mathbb{E} Y(z) - T(1 + \mathbb{E} n(z)) &= 0.
\end{align*}
\]  
(A.1)

The feasible sets of (1.20) and (A.1) are identical. To see this, fix \( w \) and choose a feasible solution \( (n(z), T(n)) \) from (1.20). Then, \( (C(z), Y(z)) = (n(z), T(n(z))) \) satisfy

\[
U(C, Y, z) := zf(C) - wC - Y
\]

\[
U(C(z), Y(z), z) \geq U(C(z'), Y(z'), z) \quad \forall z, z'
\]

To see this, pick arbitrary \( z \) and \( z' \in [z_{\min}, z_{\max}] \). Then,

\[
zf(C(z)) - wC(z) - Y(z) = zf(n(z)) - wn(z) - T(n(z)) \geq zf(n(z')) - wn(z') - T(n(z')) = zf(C(z')) - wC(z') - Y(z').
\]

The opposite is also true. Given a \( (C(z), Y(z)) \) that satisfies the truth-telling condition
(A.1), the SDP can be constructed by

\[ T(n) = \inf_T \{ \bar{Y} : U(C(z),Y(z),z) \geq U(n,\bar{Y},z) \forall z \} \,.
\]

By the truth-telling condition,

\[ U(C(z),Y(z),z) = U(C(z),T(C(z)),z) \geq U(n,T(n),z) \forall z,n. \]

Hence, agents facing \( T(\cdot) \) chooses \( n(z) = C(z) \).

Next, I reformulate the truth-telling condition into the envelope condition and the monotonicity constraint. Note that the utility function \( U(C,Y,z) \) satisfies Spence-Mirrlees condition

\[ MRS(C,Y,z) = -\frac{U_Y}{U_C} = \frac{1}{zf'(C) - w}, \quad MRS_z = -\frac{f'(C)}{(zf'(C) - w)^2} < 0. \]

Hence, the truth-telling condition is equivalent to

\[ \pi(z) := U(C(z),Y(z),z) \]

\[ \pi'(z) = U_z(C(z),Y(z),z), \quad C'(z) \geq 0 \]
By substituting $Y$ out, the problem becomes

$$\max_{w, n(z), \pi(z)} w \text{ s.t.} \begin{cases} \mathbb{E}u(\pi(z)) - u(w) = 0 \\ \pi'(z) = f(n(z)) \\ n'(z) \geq 0 \\ \mathbb{E}[zf(n(z)) - wn(z) - \pi(z)] - T(1 + \mathbb{E}n(z)) = 0. \end{cases} \quad (A.2)$$

### A.1.4 Proof of proposition 1.5

To solve (A.2) using optimal control theory, rewrite the integral constraints by

$$B'(z) = \{zf(n(z)) - wn(z) - \pi(z) - T(1 + n(z))\} g(z), \quad B(z_{\text{min}}) = B(z_{\text{max}}) = 0.$$  

$$I'(z) = (u(\pi(z)) - u(w)) g(z), \quad I(z_{\text{min}}) = I(z_{\text{max}}) = 0.$$  

Wage is constant, so $w'(z) = 0$. Monotonicity constraint becomes $n'(z) = L(z) \geq 0$. The only control variable is $L$, state variables are $\pi, w, B, I, n$, and the costate variables are $\mu_\pi, \mu_w, \mu_B, \mu_I, \mu_n$. Hamiltonian is

$$H = wg(z) + \mu_\pi f(n) + \mu_w x 0 + \mu_n L + \kappa L$$

$$+ \mu_B \{zf(n) - wn - \pi - T(1 + n)\} g(z) + \mu_I (u(\pi) - u(w)) g(z)$$

$$H_L = 0 = \mu_n + \kappa$$

$$H_\pi = -\mu_\pi' = (-\mu_B + \mu_I u'(\pi)) g(z)$$
\[ H_w = -\mu'_w = (1 - \mu_B n - \mu_I u' (w)) g (z) \]

\[ H_B = -\mu'_B = 0 \]

\[ H_I = -\mu'_I = 0 \]

\[ H_n = -\mu'_n = \mu_{\pi} f' (n) + \mu_B \left\{ z f' (n) - w - T \right\} g (z) \]

\[ \kappa L = 0, \quad \kappa \geq 0, \quad L \geq 0. \]

Boundary conditions are

\[ B (z_{\text{min}}) = B (z_{\text{max}}) = I (z_{\text{min}}) = I (z_{\text{max}}) = \mu_{\pi} (z_{\text{min}}) = \mu_{\pi} (z_{\text{max}}) \]

\[ = \mu_w (z_{\text{min}}) = \mu_w (z_{\text{max}}) = \mu_n (z_{\text{min}}) = \mu_n (z_{\text{max}}) = 0 \]

By reducing the system, the solution \((\pi, w, \mu_B, \mu_I, \mu_{\pi}, \mu_w, I, B, n, \mu_n)\) satisfy

\[ \pi' = f (n) \]

\[ w' = \mu'_B = \mu'_I = 0 \]

\[ \mu'_{\pi} = (\mu_B - \mu_I u' (\pi)) g (z) \]

\[ \mu'_w = (\mu_B n + \mu_I u' (w) - 1) g (z) \]

\[ B' (z) = \{ zf (n (z)) - wn (z) - \pi (z) - T (1 + n (z)) \} g (z) \]
\[ I'(z) = (u(\pi(z)) - u(w))g(z) \]

\[ \mu'_n = -\mu_{\pi}f'(n) - \mu_B \{ zf'(n) - w - T \} g(z) \]

\[ B(z_{\text{min}}) = B(z_{\text{max}}) = I(z_{\text{min}}) = I(z_{\text{max}}) = \mu_{\pi}(z_{\text{min}}) = \mu_{\pi}(z_{\text{max}}) \]

\[ = \mu_w(z_{\text{min}}) = \mu_w(z_{\text{max}}) = \mu_n(z_{\text{min}}) = \mu_n(z_{\text{max}}) = 0 \]

\[ \mu_n n' = 0, \mu_n \leq 0, n' \geq 0. \]

By assumption, \( n' > 0 \). Then, from \( \mu'_n = 0 \), \( n(z) \) is determined by

\[ zf'(n(z)) = \frac{w + T}{\frac{\mu_{\pi}(n)}{z\mu_B g(z)} + 1} \quad (A.3) \]

In this case, ODE reduced to \( y = (\pi, w, \mu_B, \mu_I, \mu_{\pi}, \mu_w, B, I) \) such that

\[ \pi' = f(n) \]

\[ w' = \mu_B' = \mu_I' = 0 \]

\[ \mu_{\pi}' = (\mu_B - \mu_I u'(\pi)) g(z) \]

\[ \mu_w' = (\mu_I u'(w) + \mu_B n - 1) g(z) \]

\[ B'(z) = \{ zf(n) - wn - \pi - T (1 + n) \} g(z) \]

\[ I'(z) = (u(\pi) - u(w))g(z) \]
\( B (z_{\text{min}}) = B (z_{\text{max}}) = I (z_{\text{min}}) = I (z_{\text{max}}) = \mu_{\pi} (z_{\text{min}}) = \mu_{\pi} (z_{\text{max}}) = \mu_{w} (z_{\text{min}}) = \mu_{w} (z_{\text{max}}) = 0 \)

To show that the medium-sized firms are distorted, it suffices to show \( \mu_B > 0 > \mu_{\pi} (z) \) due to A.3. To see this, note that from envelope theorem,

\[
\frac{\partial H}{\partial T} = -\mu_B E (1 + n (z)) .
\]

Since higher tax takes more resources, \( \frac{\partial H}{\partial T} < 0 \) so \( \mu_B > 0 \). By integrating \( \mu_{\pi}' \), the boundary condition implies

\[
\mu_B - \mu_I E u' (\pi (z)) = 0 \Rightarrow \mu_I > 0 .
\]

Now, let’s focus on \( \mu_{\pi}' \), which has to cross zero since \( \mu_{\pi} (z_{\text{min}}) = \mu_{\pi} (z_{\text{max}}) = 0 \). Since \( \pi' > 0 \), \( \mu_{\pi} \) must be convex,

\[
\frac{\partial \mu_{\pi}' (z)}{\partial z} = -\mu_I u'' (\pi (z)) \pi' (z) > 0 .
\]

Therefore, \( \mu_{\pi}' \) starts from negative and monotonically cross zeros. Hence, \( \mu_{\pi} < 0 \).

The fact that the optimal SDP increases the number of firms follows from the distortion. Note that when \( \mathcal{T} = 0 \), the laissez-faire economy \( T (n) = 0 \) is a feasible choice for the government. Therefore, the optimal SDP increases wage \( w^o > w^{LF} \). For small \( \mathcal{T} > 0 \), the marginal products are higher than the laissez-faire wage

\[
z f' (n (z)) = \frac{w^o + \mathcal{T}}{\frac{\mu_{\pi} (z)}{z \mu_B g (z)} + 1} > w^{LF} .
\]

Hence, employment is lower at each productivity level. Since the entrepreneurs and workers have to add up to one, it implies that the number of firms increases. \( \phi^o > \phi^{LF} \).
A.1.5 Proof of proposition 1.6

The equilibrium is characterized by

\[ \mathbb{E}u(\pi(z, w)) = u(w) \]

\[ \pi(z, w) = \max_n zn^\alpha - wn \text{ s.t. } wn \leq \lambda zn^\alpha. \]

\[ \phi\mathbb{E}n(z, w) = 1 - \phi \]

Let \( \lambda^* = \min \{\alpha, \lambda\} \). The firm's problem gives

\[ n(z, w) = \left(\frac{\lambda^* z}{w}\right)^{\frac{1}{1-\alpha}}, \quad \pi(z, w) = (1 - \lambda^*) \left\{ z \left(\frac{\lambda^*}{w}\right)^{\alpha} \right\}^{\frac{1}{1-\alpha}} \]

The indifference condition leads to

\[ w^{LF} = (1 - \lambda^*)^{1-\alpha} (\lambda^*)^\alpha \left( \mathbb{E}z^{\frac{1}{1-\alpha}} \right)^{\frac{1-\alpha}{1-\gamma}} \]

The market clearing condition implies that the equilibrium number of firms is

\[ \phi^{LF} = \frac{1 - \lambda^*}{1 - \lambda^* + \lambda^* \mathbb{E}z^{\frac{1}{1-\alpha}} (\mathbb{E}z^{\frac{1-\gamma}{1-\alpha}})^{-\frac{1}{1-\gamma}}}. \]

This is a decreasing function of \( \lambda^* \) satisfying \( \phi^{LF} \to 1 \) as \( \lambda \to 0 \). Since the planner does not face any friction, \( \phi^P \) does not depend on \( \lambda \). Hence, for small \( \lambda \), \( \phi^P < \phi^{LF} \).
A.1.6 Proof of proposition 1.7

I first solve the laissez-faire economy and then derive the efficient allocation.

(Laissez-faire) Let $V_t(a, j, z)$ be the value function. Guess $V_t(a, j, z) = v^j_t(z) a^\rho$. For instance, the problem for entrepreneurs is

$$v^E_t(z) a^\rho = \max_{c, e'} \exp \left[ (1 - \beta) \ln c + \beta \rho \ln (r_t a^\rho - c) ight]$$

$$+ \beta \ln \left\{ e' \mathbb{E} v^E_{t+1} (z')^{1-\gamma} + (1 - e') \mathbb{E} v^W_{t+1} (z')^{1-\gamma} \right\}^{1-\gamma}$$

The maximization in terms of $c$ and coefficient matching with respect to $a$ gives

$$c = \frac{1 - \beta}{1 - \beta + \beta \rho} r_t a^\rho, \quad \rho = \frac{1 - \beta}{1 - \beta \theta} \Rightarrow \begin{cases} 
  c = (1 - \beta \theta) r_t a^\theta \\
  a' = \beta \theta r_t a^\theta
\end{cases}$$

The same structure appears in worker’s problem with $r_t$ replaced by $w_t$. Since the productivity is i.i.d., the productivity and asset distribution are independent. In the equilibrium where everyone takes the same strategy $e_{it} = \phi$ for all $t$, the market clearing condition becomes

$$\begin{cases} 
  n(z, w, k) = \left( \frac{\alpha z}{w} \right)^{1-\alpha} k = \left( \frac{\alpha z}{w} \right)^{1-\alpha} a^\theta \\
  h = a^\theta
\end{cases} \Rightarrow \left( \frac{\alpha}{w_t} \right)^{\frac{1}{1-\alpha}} \mathbb{E} z^{\frac{1}{1-\alpha}} \phi = 1 - \phi \quad (A.4)$$

so $w_t$ and $r_t$ is time independent. One can also see the value functions are also time inde-
pendent and \( v^W \) does not depend on \( z \).

\[
\ln v^E (z) = (1 - \beta) \ln (1 - \beta \theta) + \beta \theta \frac{1 - \beta}{1 - \beta \theta} \ln \beta \theta \\
+ \frac{1 - \beta}{1 - \beta \theta} \ln r(z, w) + \beta \ln \left\{ \phi \mathbb{E} v^E (z')^{1-\gamma} + (1 - \phi) (v^W)^{1-\gamma} \right\}^{1-\gamma} \quad (A.5)
\]

\[
\ln v^W = (1 - \beta) \ln (1 - \beta \theta) + \beta \theta \frac{1 - \beta}{1 - \beta \theta} \ln \beta \theta \\
+ \frac{1 - \beta}{1 - \beta \theta} \ln w + \beta \ln \left\{ \phi \mathbb{E} v^E (z')^{1-\gamma} + (1 - \phi) (v^W)^{1-\gamma} \right\}^{1-\gamma} \quad (A.6)
\]

By subtracting one from the other, I can simplify the relationship between the two value functions

\[
v^E (z) = \left( \frac{r(z, w)}{w} \right)^{\frac{1-\beta}{1-\beta \theta}} v^W. \quad (A.7)
\]

The occupation choice gives the indifference condition

\[
\left( \mathbb{E} v^E (z')^{1-\gamma} \right)^{\frac{1}{1-\gamma}} = v^W.
\]

These two conditions lead to the equilibrium condition that the wage has to satisfy

\[
\mathbb{E} \left( \frac{r(z, w)}{w} \right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta \theta}} = 1. \quad (A.8)
\]

Since the return to capital is

\[
r(z, w) = (1 - \alpha) z^{\frac{\alpha}{1-\alpha}} w^{\frac{\alpha}{\alpha - 1}} z^{\frac{1}{1-\alpha}},
\]
the equilibrium wage is

\[ w^m = (1 - \alpha)^{1-\alpha} \alpha^\alpha \left( \mathbb{E} z^{(1-\gamma)(1-\beta)} \right)^{(1-\alpha)(1-\beta)} \]

and the number of firms is given by the market clearing condition

\[ \phi^m = \frac{1 - \alpha}{1 - \alpha + \alpha \mathbb{E} z^{1-\alpha} \left\{ \mathbb{E} \left( z^{1-\alpha} \right)^{1-\beta} \right\}^{1-\gamma}(1-\beta)} \cdot \]

(Planner) To derive the optimal number of firms, it suffices to consider output maximization. The problem is recursive, so fix a time period \( t \) and the aggregate asset saving \( A_t := \int a_i \, di \). This is an endogenous variable, but for the purpose of deriving the optimal number of firms, I can take it as given as will be clear. Since the investment function \( f_I \) is strictly concave and \( z_{it+1} \) is not observable at \( t \), it is optimal to choose the same saving across agents.

\[ k_{it} = k_t := f_I (a_t^E), \quad h_{it} := h_t := f_I (a_t^W). \]

Conditional on the amount of capital \((k_t, h_t)\), the problem at \( t \) reduces to

\[
\max_{Y_{t+1}, n_{it+1}, \phi_t} Y_{t+1} \text{ s.t. } \begin{cases} 
\phi_t \int n_{it+1} \, di = (1 - \phi_t) \, h_t \\
Y_{t+1} = \phi_t \int z_{it+1} f (k_{t-1}, n_{it+1}) \, di 
\end{cases}.
\]

By maximizing out \( n_{it+1} \), the aggregate output can be written as

\[ Y_{t+1} = \phi_t^{1-\alpha} (1 - \phi_t)^\alpha \left( \mathbb{E} z^{1-\alpha} \right)^{1-\alpha} k_t^{1-\alpha} h_t^\alpha. \]
Hence, the planner’s problem reduces to

\[
\max_{a_t^E, a_t^W, \phi_t, k_t, h_t} Y_{t+1} \text{ s.t. } \begin{cases} 
\phi_t a_t^E + (1 - \phi_t) a_t^W = A_t \\
Y_{t+1} = \phi_t^{1-\alpha} (1 - \phi_t)^\alpha \left( \mathbb{E} z^{\frac{1}{1-\sigma}} \right)^{1-\alpha} k_t^{1-\alpha} h_t^\alpha.
\end{cases}
\]

By optimizing out \((a_t^E, a_t^W)\), the output takes the following form

\[
Y_{t+1} = \left\{ \phi_t^{1-\alpha} (1 - \phi_t)^\alpha \right\}^{1-\theta} \left( \mathbb{E} z^{\frac{1}{1-\sigma}} \right)^{1-\alpha} (1 - \alpha)^{\theta(1-\alpha)} \alpha^\theta A_t^\theta.
\]

Hence, the optimal number of firms is \(\phi_t^P = 1 - \alpha\). One can see \(\phi_t^{LF} < \phi_t^P\) iff \(V(z) > 0\) since \(\phi_t^{LF} < 1 - \alpha\) is equivalent to

\[
\mathbb{E} z^{\frac{1}{1-\sigma}} > \left\{ \mathbb{E} \left( z^{\frac{1}{1-\sigma}} \right)^{\frac{(1-\gamma)(1-\beta)}{1-\beta}} \right\}^{\frac{1-\beta\theta}{(1-\gamma)(1-\beta)}} \Leftrightarrow \gamma + (1 - \theta) \frac{\beta}{1-\beta} > 0,
\]

which is always true under the parameter assumptions.

**A.2 Derivation of Eq.(1.12) and (1.13)**

Eq.(1.12) can be derived as follows. The key is that marginal product equalization holds in the laissez-faire economy.

\[
Y = \phi \mathbb{E} [zf (n (z, w))].
\]
By taking derivative, noting \( w = w(T) \) and \( \phi = \phi(T) \) are functions of policies

\[
\partial_T Y = \mathbb{E}zf'(n(z, w)) \partial_T \phi + \phi \mathbb{E} [zf'(n(z, w)) \partial_T n(z, w)].
\]

Since the derivative is evaluated at laissez-faire, \( zf'(n(z, w)) = w \), and \( \partial_T Y \) becomes

\[
\partial_T Y = \mathbb{E}zf'(n(z, w)) \partial_T \phi + w\phi \mathbb{E} [\partial_T n(z, w)]
\]

We can simplify the second term by substituting out the derivative of the market clearing condition

\[
\partial_T \phi \mathbb{E}n(z, w) + \phi \mathbb{E} [\partial_T n(z, w)] = -\partial_T \phi.
\]

Hence, the marginal impact can be written as

\[
\partial_T Y = \partial_T \phi \mathbb{E}zf'(n) - w\partial_T \phi (1 + \mathbb{E}n) = \partial_T \phi (\mathbb{E}\pi(z, w) - w).
\]

Note that the derivation does not use the indifference condition \( \mathbb{E}u(\pi(z, w)) = u(w) \).

Similarly, Eq.(1.13) can be derived as follows. Aggregate output is

\[
Y = \mathbb{E} [zf(n(z, w)) 1_{z \geq \bar{z}}].
\]

By taking derivative, noting \( w = w(T) \) and \( \bar{z} = \bar{z}(T) \) are functions of policies,

\[
\partial_T Y = \mathbb{E} [zf'(n(z, w)) 1_{z \geq \bar{z}} \partial_T n(z, w)] - \bar{z}f(n(\bar{z}, w)) g_z(\bar{z}) \partial_T \bar{z}.
\]
Since the derivative is evaluated at laissez-faire, \( zf'(n(z, w)) = w \), and \( \partial_T Y \) becomes
\[
\partial_T Y = w \mathbb{E}[1_{z \geq \bar{z}} \partial_T n(z, w)] - \bar{z}f(n(\bar{z}, w)) g_z(\bar{z}) \partial_T \bar{z}.
\]

We can simplify the first term by substituting out the derivative of the market clearing condition,
\[
\mathbb{E}[1_{z \geq \bar{z}} \partial_T n(z, w)] - n(\bar{z}, w) g(\bar{z}) \partial_T \bar{z} = g(\bar{z}) \partial_T \bar{z}.
\]

Hence, the marginal impact can be written as
\[
\partial_T Y = w(1 + n(\bar{z}, w)) g_z(\bar{z}) \partial_T \bar{z} - \bar{z}f(n(\bar{z}, w)) g_z(\bar{z}) \partial_T \bar{z}
\]
\[
= (\pi(\bar{z}, w) - w) (-g_z(\bar{z}) \partial_T \bar{z}).
\]

Note that the derivation does not use the indifference condition \( \pi(\bar{z}, w) = w \).

A.3 Non-pecuniary SDP

If the SDP is non-pecuniary and does not generate tax revenue, the fair comparison is (1) the economy with non-pecuniary SDP, i.e.
\[
\mathbb{E} u(\pi^{SDP}(z, w^{SDP})) = u(w^{SDP})
\]
\[
\left[ \pi^{SDP}(z, w^{SDP}), n^{SDP}(z, w^{SDP}) \right] = \max_n \begin{cases} 
z f(n) - w^{SDP} n & n \leq N \\
z f(n) - w^{SDP} \tau n - F & n > N 
\end{cases}
\]

145
and (2) the laissez-faire economy with the same amount of non-distortionary regulation costs

\[
\mathbb{E}u \left( (1 - t) \pi^{LF} (z, w^{LF}) \right) = u \left( (1 - t) w^{LF} \right)
\]

\[
\left[ \pi^{LF} (z, w^{LF}), n^{LF} (z, w^{LF}) \right] = \max_n z f (n) - w^{LF} n
\]

\[
\phi^{LF} \mathbb{E} n^{LF} (z, w^{LF}) = 1 - \phi^{LF}
\]

where \( t \) is chosen to equate the tax revenue in the first economy.

With this specification, the policy implication remains the same.

\[
\ln \left\{ (1 - t) w^{LF} \right\} - \ln w^{SDP} = -1.27\% < 0.
\]

### A.4 MLE estimation

The effective tax rates \((\tau, F)\) are estimated from the firm size distribution. Given the labor demand \( n (z, w; \tau, F) \) and the the power law density of productivity \( z \), the model-implied firm size distribution \( G_n (n; \tau, F) := P \left( n \left( \frac{z}{w}; \tau, \frac{F}{w} \right) \leq n \right) \) follows a broken power law and can be fit with the empirical firm size distribution using the maximum likelihood method. The indifference condition (1.8) serves as a constraint that restricts parameter values. The only difference between the two models is whether the indifference condition reflects entrepreneurial risks or not. I show that the indifference condition plays little part in the estimation of the effective tax rates.
Following Garicano et al. [2016], the likelihood function is constructed as follows. Given the parametric assumptions \( f(n) = n^\alpha \), the demand function can be written as

\[
n(z, w; \tau, F) = \begin{cases} 
\left( \frac{z^\alpha}{w} \right)^{\frac{1}{1-\alpha}} & z_{\min} \leq z \leq \bar{z} \\
N & \bar{z} \leq z \leq \bar{z} \\
\left( \frac{z^\alpha}{w^{\tau}} \right)^{\frac{1}{1-\alpha}} & \bar{z} < z \leq z_{\max} 
\end{cases}
\]

\[
N = \left( \frac{z^\alpha}{w} \right)^{\frac{1}{1-\alpha}} \Leftrightarrow \frac{\bar{z}}{w} = \frac{N^{1-\alpha}}{\alpha}
\]

Given the density of the productivity \( g_z(z) \), I can derive the distribution function of \( n_i = n(z_i, w) \) by using the change of variable.

\[
G_n(x) = P(n \leq x) = \begin{cases} 
0 & x < n_{\min} \\
n_{\min} - \frac{n_{\min}^{1-\beta} - n_{\max}^{1-\beta}}{n_{\min}^{1-\beta} - T n_{\max}^{1-\beta}} & n_{\min} \leq x < N \\
N - \frac{n_{\min}^{1-\beta} - n_{\max}^{1-\beta}}{n_{\min}^{1-\beta} - T n_{\max}^{1-\beta}} & N \leq x \leq \bar{n} \\
\bar{n} - \frac{n_{\min}^{1-\beta} - n_{\max}^{1-\beta}}{n_{\min}^{1-\beta} - T n_{\max}^{1-\beta}} & \bar{n} \leq x \leq n_{\max} \\
1 & n_{\max} < x
\end{cases}
\]

\[
n_{\min} = \left( \frac{z_{\min}^\alpha}{w} \right)^{\frac{1}{1-\alpha}} \\
n_{\max} = \left( \frac{z_{\max}^\alpha}{w^{\tau}} \right)^{\frac{1}{1-\alpha}} \\
1 - \beta = (1 - \alpha) \left( 1 - \beta_z \right) \\
T = \tau^{1-\beta_z}
\]

To reconcile the model in which there are no firms over the region \( N \leq n \leq \bar{n} \) with the actual data (Fig.1.1), the observed firm size distribution is assumed to be generated with
measurement error, i.e., $\tilde{n}_i = n(z_i, w) e^{-\sigma \epsilon_i}, \epsilon_i \sim N(0, 1)$. The distribution of $\tilde{n}_i$ follows

$$G_{\tilde{n}}(\tilde{n}) = P(ne^{-\sigma \epsilon} \leq \tilde{n}) = E[P(n \leq \tilde{n}e^{\sigma \epsilon} | \epsilon)]$$

$$= \frac{1}{n_{\min}^{1-\beta} - Tn_{\max}^{1-\beta}} \left[ n_{\min}^{1-\beta} \Phi \left( \frac{\ln \tilde{n}}{\sigma} \right) - Tn_{\max}^{1-\beta} \Phi \left( \frac{\ln \tilde{n}}{\sigma} \right) \right]$$

$$- \tilde{n}^{1-\beta} e^{\frac{\sigma^2(1-\beta)^2}{2}} \left\{ \Phi \left( (1 - \beta) \sigma - \frac{\ln n_{\min}}{\sigma} \right) - \Phi \left( (1 - \beta) \sigma - \frac{\ln n_{\max}}{\sigma} \right) \right\}$$

$$- T\tilde{n}^{1-\beta} \left\{ \Phi \left( \frac{\ln \tilde{n}}{\sigma} \right) - \Phi \left( \frac{\ln n_{\max}}{\sigma} \right) \right\}$$

$$- T\tilde{n}^{1-\beta} e^{\frac{\sigma^2(1-\beta)^2}{2}} \left\{ \Phi \left( (1 - \beta) \sigma - \frac{\ln n_{\max}}{\sigma} \right) - \Phi \left( (1 - \beta) \sigma - \frac{\ln n_{\min}}{\sigma} \right) \right\}$$

In addition, to utilize the bunching around $N$ and avoid the deviation of the model from the data at the extremes, the sample $\{\tilde{n}_i\}_i$ is truncated between $n_{\min} = 10$ to $n_{\max} = 10,000$. Hence, the likelihood function of the parameters $\theta = (\alpha, n_{\min}, n_{\max}, \bar{n}, \beta, \sigma, T)$ is

$$L(\theta; \{\tilde{n}_i\}_i) = \frac{\prod_i G_{\tilde{n}}(\tilde{n}_i; \theta)}{G_{\tilde{n}}(n_{\max}; \theta) - G_{\tilde{n}}(n_{\min}; \theta)} \quad (A.9)$$

The parameters are subject to the constraint specified by the indifference condition. Given the CRRA utility parameterization $u(c) = c^{1-\gamma} / (1-\gamma)$, the indifference condition is

$$\mathbb{E} \left[ \left( \frac{z}{w} f(n) - n - \left( \tau - 1 + \frac{F}{w} \right) 1_{n>N} \right)^{1-\gamma} \right] = 1 \quad (A.10)$$

Note that since the likelihood is based on truncation, the extreme values $(n_{\min}, n_{\max})$ cannot be estimated accurately. Garicano et al. [2016] avoids this problem by fixing $(\alpha, n_{\max}) = (.8, \infty)$ and maximizing the likelihood $L$ subject to their version of the indifference condition.
In this way, the indifference condition can be used to obtain $n_{\min} = \frac{\alpha}{1-\alpha}$, so the problem becomes unconstrained maximization of the likelihood (A.9) over $(\tilde{n}, \beta, \sigma, T)$. Similarly, I can fix $(\alpha, n_{\max}) = (.8, \tilde{n}_{\max})$ and maximize the likelihood (A.9) subject to the indifference condition (A.10). The estimates under these two methods are listed in the following table.

<table>
<thead>
<tr>
<th>Table A.1: Comparison</th>
<th>Baseline</th>
<th>GLV with Amadeus</th>
<th>GLV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.848</td>
<td>1.849</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.028)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>$\tilde{n}$</td>
<td>57.903</td>
<td>57.839</td>
<td>59.271</td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td>(0.901)</td>
<td>(2.051)</td>
</tr>
<tr>
<td>$\tau^{-1}$</td>
<td>0.021</td>
<td>0.021</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.1</td>
<td>0.099</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Number of Firms</td>
<td>62549</td>
<td>62549</td>
<td>41067</td>
</tr>
</tbody>
</table>

The first two columns use the same data from Amadeus 2006. The first column assumes entrepreneurial risks and the second does not. For the purpose of comparison, the third column lists the estimates from Garicano et al. [2016] that uses FICUS data.

One can see that the first two columns are almost identical. This is not surprising since the only difference between the two estimates is the shape of the indifference condition. Although the indifference condition affects $n_{\min}$, since the likelihood is truncated, it does not contain much information for the estimation of $(\tilde{n}, \beta, \sigma, T)$. In fact, if there is no observation error $\sigma = 0$, the truncation makes the likelihood estimation with and without entrepreneurial risks exactly identical. Therefore, in terms of the estimation, the existence of the entrepreneurial risks does not matter.

One can also see that the second and the third columns are not that different. Therefore, Amadeus 2006 is a good approximation of the administrative FICUS data. Hence, even if I
use FICUS data to estimate all the model parameters, I will get almost the same values as Garicano et al. [2016].

A.5 Intermediate signal

In this section, I show one way to introduce intermediate signal. The main finding is that the efficiency gain might not be monotonic over riskiness.

To study the intermediate signal, I need to take stance on two specifications, the welfare criterion and the parameterization of the signal structure. For the welfare criterion, since the model becomes heterogeneous agents whenever there are informative signals, I focus on the aggregate output instead of focusing on a particular Pareto weight. For the parametric assumption on the signal, I assume each individual observes

\[ s_i = 1_{\pi(z_i, w) \geq \eta w}, \; \eta \in [0, 1]. \]  \tag{A.11}

I could choose other often-used signal structures such as log normal, but this signal structure has several advantages in the current setting.

First, the parameter \( \eta \) has a natural interpretation. Agents who observe \( s_i = 1 \) know that they can earn at least \( \eta \% \) of their salary if they are engaged in entrepreneurship, and vice versa. The interpretation makes it clear that \( \eta = 0 \) corresponds to full-risk case, and \( \eta = 1 \) to no-risk case from the point of view of occupation choice.\(^1\) Indeed, when \( \eta = 0 \), the signal is uninformative, so the equilibrium is characterized by (1.8), while when \( \eta = 1 \), the wage

\(^1\)Agents still are uncertain about the productivity \( z_i \), but they are sure which occupation is better when \( \eta = 1 \). See the discussion right after definition 1.2.
and signal threshold satisfying (1.7) is an equilibrium. Therefore, \( \eta \in [0,1] \) is a legitimate parameter that connects the full-risk and no-risk cases.

Second, the signal structure (A.11) preserves power law. Since the productivity \( z \) follows a power law and the signal gives a truncation

\[
\pi(z_i, w) \geq \eta w \Leftrightarrow z_i \geq z(\eta, w),
\]

the distribution of the productivity conditional on the signal is still a power law. This allows me to use the same parameter values as in 1.4.1.

Given the signal structure, the equilibrium system depends on \( \eta \). For small \( \eta \), there are plenty of agents observing \( s_i = 1 \), so the equilibrium \((w, \phi)\) is determined by

\[
\mathbb{E}[u(\pi(z_i, w))|s_i = 1] = u(w), \quad \phi \mathbb{E}[n(z_i, w)|s_i = 1] = 1 - \phi, \quad P(s_i = 1) > \phi.
\]

Mathematically, the third condition is slack for small \( \eta \). As \( \eta \) increases, the third condition becomes tight, and when it binds, the indifference condition has to be relaxed. Hence, for large \( \eta \),

\[
\mathbb{E}[u(\pi(z_i, w))|s_i = 1] > u(w), \quad \phi \mathbb{E}[n(z_i, w)|s_i = 1] = 1 - \phi, \quad P(s_i = 1) = \phi.
\]

The results are shown in Fig.A.1. For each \( \eta \), the wage is normalized to \( w^{SDP} = 1 \). At the two extremes, the efficiency gain from removing the SDP is \(-1.5\%\) and \(.02\%\). For low value of \( \eta \), neither economy reacts to the increase in \( \eta \). This is because \( z_{min} > 0 \), so for the small value of \( \eta \), the information about the productivity is not informative \( \pi(z_{min}, w) \geq \eta w \).
As $\eta$ crosses some threshold, the output of both economy starts to go up since the average entrepreneurs are more productive. However, the speed of the output growth might be different, so the difference might not be monotone. In the current setting, $\eta = 45\%$ is the threshold above which the laissez-faire economy produces higher output. For different signal structure, the efficiency gain can be more smooth.

### A.6 Constrained efficiency

#### A.6.1 Constrained efficiency with intermediate signal

The constrained planner’s problem is

$$\max_{\bar{s}, w} \int_{\bar{s}} \mathbb{E}[u(\pi(z, w)) | s] dG_s + G_s(\bar{s}) u(w) \quad s.t. \quad \int_{\bar{s}} \mathbb{E}[n(z, w) | s] dG_s = G_s(\bar{s}).$$
The idea is that the planner’s choice set is restricted to the number of firms \( \phi \), or equivalently \( \bar{s} \). All other allocations, such as employment schedule and consumption, are determined by the market. This is a stronger result of efficiency and follow the spirit of Diamond [1967].

Denote the solution by \( \phi^{CP} \). By restricting the functional form, a similar result as unconstrained efficiency can be obtained.

**Claim.** Fix CRRA utility and Cobb-Douglas production functions \((u,f)\). The laissez-faire economy generates insufficient number of firms \( \phi^{LF} < \phi^{CP} \) compared to the constrained planner iff there are entrepreneurial risks.

**Proof.** The equilibrium of the laissez-faire economy is characterized by \((\bar{s}^{LF}, w^{LF})\) that satisfies

\[
\frac{1 - G_s(\bar{s})}{G_s(\bar{s})} \mathbb{E} \left[ z^{1-\alpha} \middle| s \geq \bar{s} \right] = \frac{1 - \alpha}{\alpha} \left( \mathbb{E} \left[ z^{1-\alpha} \middle| s = \bar{s} \right] \right)^{1-\gamma}.
\]

The planner solve

\[
\max_{\bar{s}} \frac{1 - G_s(\bar{s})}{1 - \gamma} \left( 1 - \alpha \right) \alpha ^\frac{\alpha}{\alpha - 1} w(\bar{s})^\frac{\alpha}{\alpha - 1} \mathbb{E} \left[ z^{1-\alpha} \middle| s \geq \bar{s} \right] + \frac{G_s(\bar{s})}{1 - \gamma} w(\bar{s})^{1-\gamma}
\]

s.t. \( w(\bar{s}) = \alpha \left( \frac{1 - G_s(\bar{s})}{G_s(\bar{s})} \mathbb{E} \left[ z^{1-\alpha} \middle| s \geq \bar{s} \right] \right)^{1-\alpha} \)

Substituting out the constraint with respect to wage, the planner’s problem becomes

\[
\max_{\bar{s}} \frac{1}{1 - \gamma} \left[ (1 - \alpha)^{1-\gamma} \mathbb{E} \left[ z^{1-\alpha} 1_{s \geq \bar{s}} \right] \left( \frac{\mathbb{E} \left[ z^{1-\alpha} 1_{s \geq \bar{s}} \right]}{G_s(\bar{s})} \right)^{-\alpha(1-\gamma)} \right]^{(1-\alpha)(1-\gamma)}
\]

153
Taking derivative with respect to $\bar{s}$ and evaluating the FOC at $(\bar{s}^{LF}, w^{LF})$ gives

$$D(\bar{s}^{LF}, w^{LF}) = \left\{ \int_{\bar{s}^{LF}}^{\infty} \mathbb{E}_{\pi | s} \left[ \frac{\partial u \left( \pi(z, w^{LF}) \right)}{\partial w} \right] dG_s + G_s(\bar{s}^{LF}) u'(w^{LF}) \right\} w'(\bar{s}^{LF})$$

$$= \frac{G_s(\bar{s}^{LF})}{1 - G_s(\bar{s}^{LF})} - (1 - \alpha)^{-\gamma} \alpha \frac{1 - \alpha \gamma}{1 - \alpha} \left( w^{LF} \right)^{1 - \gamma} \mathbb{E} \left[ z^{1 - \alpha} | s \geq \bar{s}^{LF} \right]$$

$$=: D(\bar{s}^{LF}, w^{LF})$$

$$\times \left( w^{LF} \right)^{-\gamma} (1 - G_s(\bar{s}^{LF})) u'(\bar{s}^{LF}) .$$

If $D(\bar{s}^{LF}, w^{LF}) < 0$, it means that decreasing $\bar{s}$ increases welfare, so $\phi^{LF} < \phi^{CP}$. Since $w'(\bar{s}^{LF}) < 0$, it suffices to show $\mathcal{D}(\bar{s}^{LF}, w^{LF}) > 0$. By applying the indifference condition and the market clearing condition,

$$\mathcal{D}(\bar{s}, w) = \frac{G_s(\bar{s})}{1 - G_s(\bar{s})} - \frac{\alpha}{1 - \alpha} \left( \mathbb{E} \left[ z^{1 - \alpha} | s \geq \bar{s} \right] - \mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right] \right)$$

$$= \frac{\alpha}{1 - \alpha} \left( \frac{\mathbb{E} \left[ z^{1 - \alpha} | s \geq \bar{s} \right]}{\left( \mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right] \right)^{1 - \gamma}} - \frac{\mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right]}{\left( \mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right] \right)^{1 - \gamma}} \right)$$

When $\gamma \geq 1$, $\mathcal{D}(\bar{s}^{LF}, w^{LF}) > 0$. To see this, note that, by lemma A.1, the first term is larger than 1.

$$\frac{\mathbb{E} \left[ z^{1 - \alpha} | s \geq \bar{s} \right]}{\left( \mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right] \right)^{1 - \gamma}} > \frac{\mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right]}{\left( \mathbb{E} \left[ z^{1 - \alpha} | s = \bar{s} \right] \right)^{1 - \gamma}} > 1$$

The second term is smaller than 1 since $z \mapsto z^{1 - \gamma}$ is a decreasing function, so the higher the signal is the lower the integrand is. When $\gamma \in (0, 1)$, $\mathcal{D}(\bar{s}^{LF}, w^{LF}) > 0$ still holds. To see
this, note that by Jensen,
\[
\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right] \leq \left( \mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right] \right)^{1-\gamma} \Rightarrow \frac{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right]}{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]} \leq \left( \mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right] \right)^{-\gamma}
\]

\[
\left( \frac{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]}{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]} \right)^{\frac{1}{1-\gamma}} = \left( \mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right] \right)^{\frac{1}{1-\gamma} - 1}
\]

\[
\leq \left( \left( \mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right] \right)^{1-\gamma} \right)^{\frac{1}{1-\gamma} - 1} = \left( \mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right] \right)^{\gamma}
\]

Multiplying these two equations gives
\[
\left( \frac{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]}{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right]} \right)^{\frac{1}{1-\gamma}} \cdot \frac{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right]}{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]} \leq \frac{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s = \bar{s} \right]}{\mathbb{E} \left[ z^{\frac{1}{1-\alpha}} | s \geq \bar{s} \right]} \leq 1.
\]

Since all I use is Jensen's inequality, \( \phi^{LF} = \phi^{CP} \) when there are no entrepreneurial risks. \( \square \)

The functional form assumption is necessary. If we relax the assumptions to just \( u' > 0 > u'' \), it is possible to find \( f, G \) such that \( \phi^{CP} < \phi^{LF} \). For the detail, see Ando and Matsumura [2017].

### A.6.2 Constrained efficiency with financial frictions

The constrained planner solves

\[
\max_{\phi, w} \phi \mathbb{E} u (\pi (z, w)) + (1 - \phi) u (w)
\]
s.t. \[
\begin{cases}
\pi(z, w) = \max_n zn^\alpha - wn \text{ s.t. } wn \leq \lambda zn^\alpha \\
\phi E_n(z, w) = 1 - \phi
\end{cases}
\]

The idea is the same as section A.6.1. The constrained planner can only choose the number of firms and has to let all other allocations be determined by the market.

Claim. Fix CRRA utility and Cobb-Douglas production functions \((u, f)\) and arbitrary risk structure \(G_z\) with bounded support. If the financial constraints are severe \(\lambda \approx 0\), the laissez-faire economy generates an excessive number of firms \(\phi^{CP} < \phi^{LF}\).

By substituting out \(\phi\) using the market clearing condition and applying an increasing transformation, the objective function can be written as

\[
U(w) := \left\{ \phi(w) \mathbb{E}_\pi(z, w)^{1-\gamma} + (1 - \phi(w)) w^{1-\gamma} \right\}^{\frac{1}{1-\gamma}} = w \left\{ \frac{\mathbb{E}_\pi(z, w)^{1-\gamma} + \mathbb{E}_n(z, w)}{1 + \mathbb{E}_n(z, w)} \right\}^{\frac{1}{1-\gamma}}
\]

By taking derivative with respect to \(w\),

\[
\frac{\partial}{\partial w} \ln U(w) = \frac{1}{w} + \frac{1}{1 - \gamma} \left\{ \frac{\partial}{\partial w} \frac{\mathbb{E}_\pi(z, w)^{1-\gamma}}{w^{1-\gamma}} + \frac{\partial}{\partial w} \mathbb{E}_n(z, w) - \frac{\partial}{\partial w} \frac{\mathbb{E}_n(z, w)}{1 + \mathbb{E}_n(z, w)} \right\}
\]

The constrained efficient allocation is characterized by \(\frac{\partial}{\partial w} \ln U(w^{CP}) = 0\). Evaluating the FOC at \(w = w^{LF}\) gives

\[
\frac{\partial}{\partial w} \ln U(w^{LF}) = \frac{1}{w^{LF}} + \frac{1}{1 - \gamma} \left\{ \frac{1 - \gamma \frac{1}{w^{LF} \alpha - 1}}{1 + \mathbb{E}_n(z, w^{LF})} \right\} = \frac{1}{w^{LF}} \left\{ 1 - \phi^{LF} \right\}
\]

Since \(\phi^{LF} \rightarrow 1\) as \(\lambda \rightarrow 0\), the FOC at \(w = w^{LF}\) can be negative. Hence, \(w^{LF} > w^{CP}\). By the market clearing condition \(\phi^{CP} < \phi^{LF}\).
A.6.3 Constrained efficiency with capital accumulation

The formulation of the constrained efficiency problem inherits the ex-ante view of the planner in the static model. In particular, I make the following thought experiment. Suppose all agents are in a symmetric equilibrium in which they face the same lottery and the economy is in a stationary equilibrium. If, at some period \( t \), a planner shows up and offers a new job lottery, will everyone prefer it?

Such a thought experiment can be formulated as follows. Let \( \phi = \{\phi_t\} \) be the probability of becoming entrepreneurs at each period chosen by the constrained planner. This is the only choice that the planner can control. Other decisions are made by the market, i.e., each individual \( i \) takes \( \phi \) as given and maximizes the recursive welfare by making consumption/saving decision. Formally, the agent \( i \) with the state \( (a, j, z) = (a_{it-1}, j_{it-1}, z_{it}) \) solves

\[
V_t(a, j, z; \phi) = \max_{(c,a') \in B_t(a, j, z)} u^{-1}_{it} [(1 - \beta) u_I(c_{it}) + \beta u_I \left( u^{-1} (\phi_t \mathbb{E}u (V_{t+1} (a', E, z'; \phi)) + (1 - \phi_t) \mathbb{E}u (V_{t+1} (a', W, z'; \phi))) \right)] \quad (A.12)
\]

Note that the budget set reflects that the production schedule \( \{\pi(z, w, k), n(z, w, k)\} \) is determined by the market, not by the planner.

Since all the equilibrium objects are functions of the planner’s choice \( \phi \), the wage process \( w = \{w_t\} \) also has to be a function of \( \phi \). As in the static model, the labor market clearing condition

\[
\phi_{t-1} \int n_{it} di = (1 - \phi_{t-1}) \int h_{it-1} di
\]
specifies the mapping from the process of the number of the firms $\phi$ to wage $w$, denoted by $w(\phi)$. I note that this mapping is implicitly included in the value function $V_t(a, j, z; \phi)$.

In summary, the constrained planner can be defined as follows. Fix the fundamentals $\mathcal{E} = (\beta, u, u_t, f, f_t, G_z, S_0)$. The path of the number of firms $\{w_t^{CP}, \phi_t^{CP}\}$ is constrained efficient if it solves

$$
\max_{\phi, w} \int u^{-1}(\phi_t E u(V_t(a_{it}, E, z_{it}; \phi)) + (1 - \phi_t) E u(V_t(a_{it}, W, a_{it}; \phi))) \, di
$$

subject to $\{w_t\} = w(\phi)$.

**Claim.** Fix CRRA intra-temporal utility, log inter-temporal utility, and Cobb-Douglas production and asset transformation functions. Then, the laissez-faire economy generates an insufficient number of firms $\phi^{LF} < \phi^{CP}$.

**Proof.** The Bellman equations of the constrained planner’s problem after maximizing out consumption-saving decision are the same, i.e., Eq. (A.5), (A.6) and (A.7) all hold. However, the indifference condition no longer holds. Still, I can derive the constrained planner’s objective function. Note that by (A.7), the value of the next period can be written as

$$
\left\{ \phi E v^E(z)^{1-\gamma} + (1 - \phi)(v^W)^{1-\gamma} \right\} \frac{1}{1-\gamma} = \left\{ \phi E \left( \frac{r(z, w)}{w} \right)^{(1-\beta)(1-\gamma)} + 1 - \phi \right\} \frac{1}{1-\gamma} v^W. \quad (A.13)
$$

By substituting this condition into the worker’s Bellman equation (A.6), I can solve the
worker’s value function $v^E$ in closed form

$$\ln v^W = \ln (1 - \beta \theta) + \frac{\beta \theta}{1 - \beta \theta} \ln \beta \theta + \frac{1}{1 - \beta \theta} \ln w$$

$$+ \frac{\beta}{1 - \beta} \ln \left\{ \phi E \left( \frac{r(z,w)}{w} \right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta \theta}} + 1 - \phi \right\} \frac{1}{1-\gamma}$$  \hspace{1cm} (A.14)

As a result, the planner maximizes (A.13) subject to (A.14) and the market clearing condition (A.4). Eq.(A.13) is equivalent to the objective function because the pre-determined asset distribution and the current productivity are independent $a_{it} \perp z_{it}$ and the value functions are multiplicative with respect to asset $V_t(a,j,z) = v^j_t(z)a^\rho$. In other words, the constrained planner solves

$$\max_{\phi,w} \left\{ \phi E \left( \frac{r(z,w)}{w} \right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta \theta}} + 1 - \phi \right\}^{\frac{1}{1-\gamma}} v^W$$

s.t.

$$\frac{1}{1-\alpha} E z^\frac{1}{1-\alpha} w^{\frac{1}{1-\alpha}} - 1 - \phi = 1 - \phi$$

(A.14)

By taking log of the objective function, the FOC is

$$U'(w) w = \frac{1}{1 - \beta \theta} + \frac{1}{(1 - \gamma)(1 - \beta)(1 - \alpha)} \left\{ 1 - \phi - \frac{(1-\beta)(1-\gamma)}{1-\beta \theta} E \left( \frac{r(z)}{w} \right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta \theta}} + \frac{1}{\phi - 1} \right\}.$$  \hspace{1cm} (A.15)

Since the indifference condition $E \left( \frac{r(z)}{w} \right)^{\frac{(1-\beta)(1-\gamma)}{1-\beta \theta}} = 1$ holds in the equilibrium of the laissez-faire economy, (see Eq.(A.8))

$$U'(w^{LF}) w^{LF} = \frac{1}{1 - \beta \theta} \left( 1 - \phi^{LF} \right).$$
Hence, $\phi^{CP} \geq \phi^{LF}$ if and only if $\phi^{LF} \leq 1 - \alpha$, which is then equivalent to $\gamma + (1 - \theta) \frac{\beta}{1 - \beta} \geq 0$.

Since the inequality is always positive, the laissez-faire economy always generates insufficient firms. \qed
Appendix B

Appendix to Chapter 2

B.1 Proofs of propositions

B.1.1 Proof of proposition 2.3

The proof is divided into two parts. The first part shows the concavity of \( U(\phi) \). The second part proves \( \phi^m < \phi^p \).

B.1.1.1 Concavity of \( U(\phi) \)

Note that the market clearing condition implies

\[
\pi(z, w(\phi)) = (1 - \alpha) \alpha^{1-\alpha} w(\phi) \frac{\alpha}{1-\alpha} z^{\frac{1}{1-\alpha}} = (1 - \alpha) \left( E_z z^{\frac{1}{1-\alpha}} \right)^{-\alpha} \left( \frac{\phi}{1-\phi} \right)^{-\alpha} z^{\frac{1}{1-\alpha}}.
\]
Hence, $U(\phi)$ can be written as

$$U(\phi) = \phi \mathbb{E} u(\pi(z, w(\phi))) + (1 - \phi) u(w(\phi))$$

$$= \frac{1}{1 - \gamma} \left\{ \theta_e \phi^{1-\alpha(1-\gamma)} (1 - \phi)^{\alpha(1-\gamma)} + \theta_w \phi^{(1-\alpha)(1-\gamma)} (1 - \phi)^{1-(1-\alpha)(1-\gamma)} \right\} - 1$$

where $\theta_e > 0$ and $\theta_w > 0$ are independent of $\phi$

$$\theta_e = (1 - \alpha)^{1-\gamma} \left( \mathbb{E}_z \frac{1}{1-a} \right)^{\alpha(\gamma-1)} \mathbb{E}_z \frac{1}{1-a}, \quad \theta_w = \left( \mathbb{E}_z \frac{1}{1-a} \right)^{(1-\alpha)(1-\gamma)} \alpha^{1-\gamma}.$$ 

Note that, for arbitrary $a \in \mathbb{R}$ and $b \in \mathbb{R}$

$$\frac{d}{d\phi} \phi^a (1 - \phi)^b = \phi^a (1 - \phi)^b \left\{ \frac{a}{\phi} - \frac{b}{1 - \phi} \right\} = \phi^{a-1} (1 - \phi)^{b-1} \left\{ a (1 - \phi) - b \phi \right\}$$

$$\frac{d^2}{d\phi^2} \phi^a (1 - \phi)^b = \phi^a (1 - \phi)^b \left\{ \frac{a (a - 1)}{\phi^2} + \frac{b (b - 1)}{(1 - \phi)^2} \right\} = \frac{2ab}{\phi (1 - \phi)}.$$ 

The curly bracket of $U$ has two such terms, one for $(a, b) = (1 - \alpha (1 - \gamma), \alpha (1 - \gamma))$ and another for $(a, b) = ((1 - \alpha) (1 - \gamma), 1 - (1 - \alpha) (1 - \gamma))$. In either case, the second order derivative is negative

$$\frac{a (a - 1)}{1 - \gamma} = \frac{b (b - 1)}{1 - \gamma} = -ab = -(1 - \alpha + \alpha \gamma) < 0$$

$$\frac{a (a - 1)}{1 - \gamma} = \frac{b (b - 1)}{1 - \gamma} = -ab = -(1 - \alpha) \{ \alpha + (1 - \alpha) \gamma \} < 0.$$ 

Hence, $U''(\phi) < 0$ for all $\phi \in (0, 1)$ and the problem is strictly concave. To show the existence and uniqueness, it suffices to show that the first order condition ranges from negative to
positive. Indeed, it is tedious but straightforward to show the signs of $\lim_{\phi \to 0} U'(\phi)$ and $\lim_{\phi \to 1} U'(\phi)$ are opposite.

\[
U'(\phi) = \frac{\theta_e}{1-\gamma} \phi^{-\alpha(1-\gamma)} (1-\phi)^{\alpha(1-\gamma)-1} \{1-\phi-\alpha(1-\gamma)\} + \frac{\theta_w}{1-\gamma} \phi^{(1-\alpha)(1-\gamma)-1} (1-\phi)^{-(1-\alpha)(1-\gamma)} \{(1-\alpha)(1-\gamma) - \phi\}
\]

If $\gamma > 1$, $\lim_{\phi \to 0} U'(\phi) = \infty > 0 = -\infty = \lim_{\phi \to 1} U'(\phi)$. If $0 < \gamma < 1$, $\lim_{\phi \to 1} U'(\phi) = \infty > 0 > -\infty = \lim_{\phi \to 0} U'(\phi)$. Finally, if $\gamma = 1$, the utility function becomes log, and

\[
U'(\phi) = \ln \left( \frac{1-\alpha}{\alpha} \left( \frac{Ez^{1-\alpha}}{1-\alpha} \right)^{-1} \right) + \frac{E \ln z}{1-\alpha} + \ln (1-\phi) - \ln \phi - \frac{\alpha}{1-\phi} + \frac{1-\alpha}{\phi}.
\]

By using the L’Hospital’s rule, $\lim_{\phi \to 0} U'(\phi) = \infty > 0 > -\infty = \lim_{\phi \to 1} U'(\phi)$. Hence, $(w^{cp}, \phi^{cp}) \in \mathbb{R}^+ \times (0,1)$.

\subsection{Proof of $\phi^m < \phi^{cp}$}

As the discussion after the theorem, the FOC of the planner evaluated at the market allocation is

\[
U'(\phi^m) = \left( \phi^m \frac{\partial \mathbb{E}u(\pi(z,w^m))}{\partial w} + (1-\phi^m) \frac{\partial u(w^m)}{\partial w} \right) w'(\phi^m).
\]

By taking derivative of the definition of risk premium,

\[
\frac{\partial \mathbb{E}u(\pi(z,w^m))}{\partial w} = u'(\mathbb{E}\pi(z,w^m) - R(w^m)) \left\{ \frac{\partial \mathbb{E}\pi(z,w^m)}{\partial w} - R'(w^m) \right\}.
\]
Note that in market equilibrium $\mathbb{E}_\pi (z, w^m) - R (w^m) = w^m$, and by envelope theorem and market clearing condition

$$\frac{\partial \mathbb{E}_\pi (z, w^m)}{\partial w} = -\mathbb{E}_n (z, w^m) = -\frac{1 - \phi^m}{\phi^m}.$$ 

Hence, $U' (\phi^m)$ can be written as

$$U' (\phi^m) = -\phi^m u' (w^m) w' (\phi^m) R' (w^m).$$

If $(u, f)$ is CRRA & CD, the indifference condition and market clearing condition lead to

$$\phi^m := \frac{1 - \alpha}{1 - \alpha + \alpha \mathbb{E} z^{\frac{1}{1-\alpha}} \left( \mathbb{E} z^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}}}. $$

Let’s use the risk premium to prove the result,

$$\begin{align*}
\mathbb{E} u (\pi (z, w)) &= \left( \frac{(1-\alpha) \alpha^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}}} {1-\gamma} \right) \mathbb{E} z^{\frac{1-\gamma}{1-\alpha} - 1} \\
\mathbb{E} \pi (z, w) &= (1 - \alpha) \alpha^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \mathbb{E} z^{\frac{1}{1-\alpha}} \\
\Rightarrow R (w) &= (1 - \alpha) \alpha^{\frac{1}{1-\alpha}} w^{\frac{\alpha}{1-\alpha}} \left\{ \mathbb{E} z^{\frac{1}{1-\alpha}} - \left( \mathbb{E} z^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}} \right\}
\end{align*}$$

Note that by Jensen’s inequality in lemma A.1, the curly bracket term is positive. By taking derivative of $R (w)$,

$$R' (w) = -\alpha^{\frac{1}{1-\alpha}} w^{\frac{1}{1-\alpha}} \left\{ \mathbb{E} z^{\frac{1}{1-\alpha}} - \left( \mathbb{E} z^{\frac{1-\gamma}{1-\alpha}} \right)^{\frac{1}{1-\gamma}} \right\} < 0.$$
Since $R'(w) < 0$ for all $w$, it is also true for $w = w^m$ and therefore $U'(\phi^m) > 0$. Combined with the concavity of $U' (\phi), \phi^m < \phi^{cp}$.

**B.1.2 Proof of proposition 2.4.**

For the third statement, we fix $\alpha \in (0, 1)$ and construct $(u, G)$. Define a strictly concave utility function $u$ by

$$u(c) = \begin{cases} 
\frac{c^{1-\gamma} - 1}{1-\gamma} & c \leq 1 \\
\frac{c^{1-\epsilon} - 1}{1-\epsilon} & c \geq 1
\end{cases}, \quad u'(c) = \begin{cases} 
e^{-\gamma} & c \leq 1 \\
\epsilon^{-\epsilon} & c \geq 1
\end{cases}, \quad u''(c) = \begin{cases} 
-\epsilon^\gamma & c \leq 1 \\
-\epsilon & c \geq 1
\end{cases}.$$ 

This function satisfies $u' > 0 > u''$ and Inada conditions. We choose $G$ such that the market equilibrium is $w^m = 1$. In particular, $z$ is Bernoulli taking $z_1 > 0$ with probability $p \in (0, 1)$ and $z_2 > z_1$ with probability $1 - p$ where $(z_1, z_2, p)$ satisfy $\mathbb{E} u (\pi(z, 1)) = u(1)$, or

$$\frac{\left(1 - \alpha\right) \alpha^{\frac{1}{1-\alpha}} z_1^{\frac{1}{1-\alpha}}}{1 - \gamma} - 1 = \frac{\left(1 - \alpha\right) \alpha^{\frac{1}{1-\alpha}} z_2^{\frac{1}{1-\alpha}}}{1 - \epsilon} - 1 \quad (1 - p) = 0.$$ 

The risk premium $R'(w)$ evaluated at $w = w^m = 1$ satisfy

$$\left(1 - \alpha\right) \alpha^{\frac{1}{1-\alpha}} z_1^{\frac{1}{1-\alpha}} \frac{\alpha p}{\alpha - 1} + \left(1 - \alpha\right) \alpha^{\frac{1}{1-\alpha}} z_2^{\frac{1}{1-\alpha}} \frac{1 - \alpha(1-p)}{\alpha - 1} - R'(w^m)$$
Let $\varepsilon \to 0$ be arbitrarily close to 0. Then,

$$
R'(w^m) \approx -\frac{\alpha p}{1-\alpha} \left( (1-\alpha) \alpha^{\frac{1}{1-\alpha}} z_1^{\frac{1}{1-\alpha}} - \left( (1-\alpha) \alpha^{\frac{1}{1-\alpha}} z_1^{\frac{1}{1-\alpha}} \right)^{-\gamma} \right).
$$

Since $E \pi(z, 1) = u(1)$, $(1-\alpha) \alpha^{\frac{1}{1-\alpha}} z_1^{\frac{1}{1-\alpha}} < 1$, and therefore $R'(w^m) > 0$ and $U'(\phi^m) < 0$.

The strict concavity of $U$ can be established in the same way as the proof of proposition 2.3 by noting the concavity of each term of the objective function

$$
U(\phi) = \phi \left\{ \frac{\pi(z_1, w(\phi))^{1-\xi_1} - 1}{1-\xi_1} p + \frac{\pi(z_2, w(\phi))^{1-\xi_2} - 1}{1-\xi_2} (1-p) \right\} + (1-\phi) \frac{w(\phi)^{1-\xi_3} - 1}{1-\xi_3}
$$

$$
= \frac{\theta_1}{1-\xi_1} \phi^{1-\alpha(1-\xi_1)} (1-\phi)^{\alpha(1-\xi_1)} + \frac{\theta_2}{1-\xi_2} \phi^{1-\alpha(1-\xi_2)} (1-\phi)^{\alpha(1-\xi_2)}
$$

$$
+ \frac{\theta_3}{1-\xi_3} \phi^{(1-\alpha)(1-\xi_3)} (1-\phi)^{(1-\alpha)(1-\xi_3)} - \left( \frac{p}{1-\xi_1} + \frac{1-p}{1-\xi_2} \right) \phi - \frac{1-\phi}{1-\xi_3}
$$

where $(\xi_1, \xi_2, \xi_3)$ can take various combinations of values in $\{\gamma, \varepsilon\}$ depending on $\phi$, and $(\theta_1, \theta_2, \theta_3)$ are positive constants independent of $\phi$. Hence, $\phi^{cp} < \phi^m$ follows.

### B.2 Heuristic derivation of Eq. (2.18)

For the mathematical proof, see Appendix B.1.1. We describe the intuitive derivation.

Note that the aggregate resources consumed by the entrepreneurs $\Pi(\phi) := \phi E \pi(z, w)$ and the workers $W(\phi) = (1-\phi) w$ are constant fractions of output

$$
Y(\phi) = \left( E z^{\frac{1}{1-\alpha}} \right)^{1-\alpha} \phi^{1-\alpha} (1-\phi)^{\alpha}
$$

(B.1)
due to the Cobb-Douglas technology

\[ \Pi(\phi) = (1 - \alpha) Y(\phi), \quad W(\phi) = \alpha Y(\phi). \]

Hence, the welfare function, using the risk premium (2.17), can be written as

\[ U(\phi) = \phi u \left( \frac{(1 - \alpha) Y(\phi)}{\phi} - R(w(\phi)) \right) + (1 - \phi) u \left( \frac{W(\phi)}{1 - \phi} \right). \]

By taking derivative and evaluating it at the market equilibrium \( \phi = \phi_m \),

\[ U'(\phi_m) = u'(w_m) \phi_m \left\{ Y'(\phi^m) - \left( \frac{1 - \alpha}{\phi^m} - \frac{\alpha}{1 - \phi^m} \right) Y(\phi^m) - R'(w_m) w'(\phi^m) \right\} \]

Let’s interpret the curly bracket term. The first term is the aggregate output increase due to better occupation allocation \( Y'(\phi^m) > 0 \). This effect is canceled by the second term, representing the compensation for the marginal agent who changes occupation from worker to entrepreneur

\[ Y'(\phi) - \left( \frac{1 - \alpha}{\phi} - \frac{\alpha}{1 - \phi} \right) Y(\phi) = 0. \]

One can see this from Eq. (B.1). \(^1\) Hence, the marginal welfare can be written as

\[ U'(\phi^m) = -\phi^m u'(w^m) w'(\phi^m) R'(w^m). \]

\(^1\)Actually, the result does not depend on Cobb-Douglas specification.
Appendix C

Appendix to Chapter 3

C.1 Proof of theorem 3.1

The proof of the theorem is by comparison of coefficients. At the end, the system of coefficients boils down to a cubic polynomial of $\beta$. The second order condition for the dealer then selects the unique negative root. All other coefficients are uniquely determined once $\beta$ is obtained. Since a cubic polynomial equation has a closed form solution, all the equilibrium coefficients can be written in closed forms. With additional calculations, we can also derive welfare of agents in closed forms. We first prove existence and uniqueness. $BC\beta_1\gamma_1\beta_0\gamma_1 \neq 0$ is assumed until it is proven at the end.

Existence and uniqueness

Proof is by guessing and verifying $p(z, s) = A + B(z + Cs)$.

Traders’ problem
For each $i = I, O$, by log normality we get

$$\arg \max_x E_i [−e^{−\theta (d−p)x}] = \arg \max_x (E_i d − p) x − \frac{\theta}{2} (V_i d) x^2 = \frac{E_i d − p}{\theta V_i d}.$$ 

By the joint normality of $X$ and $p(z,s) = A + B (z + Cs)$, the moments of return $d$ conditional on the traders’ information are

$$E_I d = E [d|z] = \frac{\kappa_d \bar{d} + \kappa_\epsilon z}{\kappa_d + \kappa_\epsilon},$$

$$V_I d = V [d|z] = \frac{1}{\kappa_d + \kappa_\epsilon},$$

$$E_O d = E [d|z + Cs] = \bar{d} + \frac{\kappa_d^{-1}}{\kappa_d^{-1} + \kappa_\epsilon^{-1} + C^2 \kappa_s^{-1}} \left( \frac{p − A}{B} − Cs − \bar{d} \right),$$

$$V_O d = \kappa_d^{-1} − \frac{\kappa_d^{-2}}{\kappa_d^{-1} + \kappa_\epsilon^{-1} + C^2 \kappa_s^{-1}}.$$

Hence, insiders’ best response is

$$x_I^B (z, p) = \alpha_I + \beta_I p + \gamma_I z$$

where

$$\alpha_I = \frac{\kappa_d \bar{d}}{\theta}, \beta_I = −\frac{\kappa_d + \kappa_\epsilon}{\theta}, \gamma_I = \frac{\kappa_\epsilon}{\theta}.$$  \hspace{1cm} \text{(C.1)}$$

Outsider’s best response is

$$x_O^B (p) = \alpha_O + \beta_O p$$

where

$$\alpha_O = \frac{(\kappa_d^{-1} + C^2 \kappa_s^{-1}) \bar{d} − \kappa_d^{-1} (\frac{A}{B} + Cs)}{\theta \kappa_d^{-1} (\kappa_\epsilon^{-1} + C^2 \kappa_s^{-1})}, \beta_O = \frac{1}{\theta} \kappa_d^{-1} − \frac{(\kappa_d^{-1} + \kappa_\epsilon^{-1} + C^2 \kappa_s^{-1})}{\theta \kappa_d^{-1} (\kappa_\epsilon^{-1} + C^2 \kappa_s^{-1})}.$$  \hspace{1cm} \text{(C.2)}$$

The total demand is

$$x^B (z, p) = \alpha + \beta p + \gamma z$$

where

$$\alpha = \lambda \alpha_I + (1 − \lambda) \alpha_O, \beta = \lambda \beta_I + (1 − \lambda) \beta_O, \gamma = \lambda \gamma_I.$$  \hspace{1cm} \text{(C.3)}$$
Dealer’s problem

Given the total demand, the dealer can infer \( z \), and therefore the dealer’s problem is

\[
\arg \max \limits_p E_D \left[ -e^{-\theta_D \left\{ (s-x^B(z,p))d+p\alpha^B(z,p) \right\}} \right]
\]

\[
= \arg \max \limits_p (s - \alpha - \beta p - \gamma z) E \left[ d | z \right] + p (\alpha + \beta p + \gamma z) - \frac{\theta_D}{2} (s - \alpha - \beta p - \gamma z)^2 V \left[ d | z \right].
\]

The first order condition with respect to \( p \) gives \( p(z,s) = A + B (z + C s) \) where

\[
[A] : A = \frac{\alpha (\kappa_d + \kappa_e) - \beta (\theta_D \alpha + \kappa_d \bar{d})}{\theta_D \beta^2 - 2 \beta (\kappa_d + \kappa_e)}
\]

\[
[B] : B = \frac{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)}{\theta_D \beta^2 - 2 \beta (\kappa_d + \kappa_e)}
\]

\[
[C] : C = \frac{\theta_D \beta}{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)}.
\]

The second order condition is

\[
\beta \left( \beta - \frac{2 (\kappa_d + \kappa_e)}{\theta_D} \right) > 0.
\]

Fixed point

Recall \((\theta, \theta_D, \bar{d}, \bar{s}, \kappa_d, \kappa_s, \kappa_e, \lambda)\) is exogenous. Existence and uniqueness of an equilibrium is equivalent to those of the eleven parameters \((\alpha_I, \beta_I, \gamma_I, \alpha_O, \beta_O, \alpha, \beta, \gamma, A, B, C)\) that satisfy \((C.1), (C.2), (C.3), (C.4)\) and \((C.5)\).

We can reduce this problem to finding a root of the equation that only contains \( \beta \). To see this, note that \((\alpha_I, \beta_I, \gamma_I)\) and therefore \( \gamma \) are already functions of exogenous parameters.
By substituting (C.2) into (C.3), the problem reduces to finding \((\alpha, \beta, A, B, C)\) satisfying

\[
\alpha = \lambda \alpha_I + (1 - \lambda) \frac{(\kappa^{-1}_d + C^2\kappa^{-1}_s) \bar{d} - \kappa^{-1}_d \left( \frac{\bar{A}}{\bar{B}} + C \bar{s} \right)}{\theta \kappa^{-1}_d (\kappa^{-1}_d + C^2\kappa^{-1}_s)}
\]

\[
\beta = \lambda \beta_I + (1 - \lambda) \frac{\frac{1}{\bar{B}} \kappa^{-1}_d - \left( \kappa^{-1}_d + \kappa^{-1}_e + C^2\kappa^{-1}_s \right)}{\theta \kappa^{-1}_d (\kappa^{-1}_e + C^2\kappa^{-1}_s)}
\]

\[
[A] : A = \frac{\alpha (\kappa_d + \kappa_e) - \beta (\theta_D \alpha + \kappa_d \bar{d})}{\theta_D \beta^2 - 2 \beta (\kappa_d + \kappa_e)}
\]

\[
[B] : B = \frac{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)}{\theta_D \beta^2 - 2 \beta (\kappa_d + \kappa_e)}
\]

\[
[C] : C = \frac{\theta_D \beta}{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)}
\]

\[
[SOC] : \beta \left( \beta - 2 \frac{\kappa_d + \kappa_e}{\theta_D} \right) > 0.
\]

By substituting \([B]\) and \([C]\) into \([\beta]\), we can obtain an equation that contains only \(\beta\). Once \(\beta\) that satisfies both this equation and \([SOC]\) is obtained, \((C, B)\) can be uniquely determined by \([C]\) and \([B]\). \((\alpha, A)\) is then the unique solution of a system of linear equations, \([\alpha]\) and \([A]\).

Now we show such \(\beta\) exists uniquely. By substituting \([B]\) and \([C]\) into \([\beta]\),

\[
b(\beta) := b_0 + b_1 \beta + b_2 \beta^2 + b_3 \beta^3 = 0, \quad (C.6)
\]
where coefficients \((b_0, b_1, b_2, b_3)\) are

\[
\begin{align*}
b_0 &= \gamma^2 \kappa_s (\kappa_d + \kappa_e)^3 \\
b_1 &= -\frac{\kappa_s^2 \pi^2}{\bar{\theta}} (\kappa_d + \kappa_e)^2 (1 + 2\theta_D) \\
b_2 &= \kappa_s (\kappa_d + \kappa_e) \left\{ (\gamma \theta_D + \kappa_e)^2 - 2(\theta + (1 - \lambda) \kappa_e) (\gamma \theta_D + \kappa_e) - (1 - \lambda) \kappa_e \gamma \theta_D \right\} \cdot \\
&\quad + \kappa_e \lambda \theta_D^2 (\kappa_d + \kappa_e) + \kappa_d \kappa_e \theta_D^2 (1 - \lambda) \\
b_3 &= \kappa_s \gamma \theta_D (\gamma \theta_D + 2 \kappa_e) + \kappa_e \kappa_s \gamma \theta_D^2 (1 - \lambda) + \kappa_s \kappa_e^2 \left\{ (1 - \lambda) \theta_D + \theta \right\} + \kappa_e \theta_D^2
\end{align*}
\]

\(b_0 = b(0) > 0\) and \(b_3 > 0\) implies that there is a \(\beta < 0\) that satisfies \(b(\beta) = 0\). Since \(\beta < 0\) satisfies \([SOC]\), we obtain existence. For the uniqueness of \(\beta\), it suffices to show \(b'(0) = b_1 < 0\) and \(b'\left(\frac{2 \kappa_d + \kappa_e}{\theta_D}\right) > 0\). Indeed,

\[
\begin{align*}b\left(\frac{2 \kappa_d + \kappa_e}{\theta_D}\right) &= \frac{\kappa_e (\kappa_d + \kappa_e)^3}{\theta^2 \theta_D^3} \left\{ \kappa_s \kappa_e (2 \theta + \theta_D) (2 \theta + \lambda \theta_D)^2 + 4 \theta^2 \theta_D^3 \frac{\kappa_d + \lambda \kappa_e}{\kappa_d + \kappa_e} + 8 \theta^3 \theta_D^2 \right\} > 0 \\
b'\left(\frac{2 \kappa_d + \kappa_e}{\theta_D}\right) &= \frac{\kappa_e (\kappa_d + \kappa_e)^2}{\theta^2 \theta_D^2} \left\{ \kappa_s \kappa_e (3 \theta + 2 \theta_D) (2 \theta + \lambda \theta_D)^2 + 4 \theta^2 \theta_D^3 \frac{\kappa_d + \lambda \kappa_e}{\kappa_d + \kappa_e} + 12 \theta^2 \theta_D^2 \right\} > 0.
\end{align*}
\]

**Signs and Regions**

We next show \(\beta_I < \beta < \beta_O < 0 < \gamma < \gamma_I\), \(C < 0 < B\) and \(BC\beta_I \gamma_I \beta_O \beta \gamma \neq 0\).

\(0 < \gamma = \lambda \gamma_I < \gamma_I\) follows by \((C.1)\) and \((C.3)\). We know \(\beta < 0\) by the above argument. \(C < 0 < B\) follows from \([C]\) and \([B]\). \(\beta_I < \beta < \beta_O < 0\) comes from the fact that \(\beta\) is a convex combination of \((\beta_I, \beta_O)\) and

\[
b(\lambda \beta_I) = \frac{(1 - \lambda) \kappa_d \kappa_e \lambda^2 \theta_D^2 (\kappa_d + \kappa_e)^2}{\theta^2} > 0
\]
Finally, we show \( BC\beta \gamma_1 \beta_0 \gamma \neq 0 \). \( \gamma_1 \beta_1 \neq 0 \) follows from the closed form solutions (C.1) and (C.3). \( BC \beta \neq 0 \) is proven by contradiction next. \( \beta_0 \neq 0 \) follows from \( \beta \neq 0 \).

To show \( \beta \neq 0 \), suppose \( \beta = 0 \). Then, the dealer’s objective is

\[
\max_p (s - \alpha - \gamma z) E [d|z] + p (\alpha + \gamma z) - \frac{\theta_D}{2} (s - \alpha - \gamma z)^2 V[d|z].
\]

For equilibrium price to exist for all \( z \in \mathbb{R} \), \( \alpha = \gamma = 0 \), a contradiction to \( \gamma > 0 \).

To show \( B \neq 0 \), suppose \( B = 0 \). Then, the price function tells nothing about the signal \( z \). Then the aggregate demand becomes

\[
x^B (z, p) = \frac{\kappa_d}{\theta} \bar{d} - \frac{\kappa_d + \lambda \kappa_e}{\theta} p + \frac{\lambda \kappa_e}{\theta} z.
\]

Given this demand function, the dealer’s optimization implies

\[
\begin{align*}
B &= \frac{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)}{\theta_D \beta' - 2 \beta (\kappa_d + \kappa_e)} = 0 \\
\Rightarrow 0 &= \gamma (\kappa_d + \kappa_e) + \frac{\kappa_d + \lambda \kappa_e}{\theta} (\theta_D \gamma + \kappa_e) > 0,
\end{align*}
\]

a contradiction.
To see $C \neq 0$, suppose $C = 0$. Since $B \neq 0$, the price function fully reveals $z$.

\[
\begin{align*}
x^B_I (z, p) &= \frac{\kappa d}{\theta} \bar{d} - \frac{\kappa d + \kappa_e}{\theta} p + \frac{\rho}{\theta} z \\
x^B_O (p) &= \frac{1}{\theta} (\kappa_d \bar{d} - \frac{A}{B}) + \frac{1}{\theta} \left\{ \frac{\rho}{B} - (\kappa_d + \kappa_e) \right\} p \\
\end{align*}
\]

\[
\Rightarrow x^B (z, p) = \frac{1}{\theta} \left\{ \kappa_d \bar{d} - (1 - \lambda) \frac{A}{B} \right\} + \frac{1}{\theta} \left\{ \frac{1 - \lambda}{B} \kappa_e - (\kappa_d + \kappa_e) \right\} p + \frac{\lambda \kappa_e}{\theta} z.
\]

Given $x^B (z, p) = \alpha + \beta p + \gamma z$, $\beta \neq 0$, and $B \neq 0$, three possibilities have to be considered.

If $[SOC]$ is met, the optimal pricing is given by (C.4). Hence,

\[
\frac{\theta_D \beta}{\gamma (\kappa_d + \kappa_e) - \beta (\theta_D \gamma + \kappa_e)} = 0 \Rightarrow \beta = 0,
\]

a contradiction to $\beta \neq 0$. If $\beta \left( \beta - \frac{2(\kappa_d + \kappa_e)}{\theta_D} \right) < 0$, there is no optimal price. If $\beta \left( \beta - \frac{2(\kappa_d + \kappa_e)}{\theta_D} \right) = 0$, $\beta = 2 \frac{\kappa_d + \kappa_e}{\theta_D}$. The dealer’s objective function becomes

\[
\arg \max_p (s - \alpha - \beta p - \gamma z) E [d|z] + p (\alpha + \beta p + \gamma z) - \frac{\theta_D}{2} (s - \alpha - \beta p - \gamma z)^2 V [d|z] = \arg \max_p \{ \alpha + \gamma z + 2 (\kappa_d + \kappa_e) (s - \alpha - \gamma z - E [d|z]) \} p.
\]

There is no way for the coefficient of $p$ to be 0 identically for all $s$ and $z$, implying there is no optimal price in this case.
C.2 Proof of proposition 3.1

Since \( s - x^B_z \) is observable to the dealer, noting \( d \) being independent of \( \xi \), we get

\[ -E_D e^{-\theta \left\{ (s-x^B_z d + px^B_z)^2 \right\}} \]

\[ = -e^{-\theta px^B_z} E_D e^{-\theta (s-x^B_z)^2} E_D e^{-\theta (s-x^B_z)^2} \]

Each part can be calculated in closed form. The second term becomes

\[ E_D e^{-\theta (s-x^B_z)^2} = e^{-\theta (s-x^B_z)^2} V_D d \]

The third term is

\[ E_D e^{-\theta \frac{\alpha}{2} (s-x^B_z)^2} = e^{-\theta \frac{\alpha}{2} (s-x^B_z)^2} E_D e^{-\theta (s-x^B_z)^2} \]

\[ \xi - \theta \frac{\alpha}{2} \xi^2 \]
\[ E_D e^{-\theta \alpha (s - x^B(z, p)) \xi - \frac{\alpha^2}{2} \xi^2} \]

\[ = \frac{1}{\sqrt{2\pi \sigma^2}} \int e^{-\theta \alpha (s - x^B(z, p)) \xi - \frac{\alpha^2}{2} \xi^2} e^{-\frac{\xi^2}{2\sigma^2}} d\xi \]

\[ = \exp \left( \frac{1}{2} \left( \frac{\theta \alpha}{\sigma^2} + \frac{1}{\sigma^2} + \theta \alpha \right) \xi - \theta \alpha^2 \right) \frac{\sqrt{2\pi} \left( \frac{1}{\sigma^2} + \theta \alpha \right)^{-1}}{\sqrt{2\pi \sigma^2}} \]

\[ \times \frac{1}{\sqrt{2\pi} \left( \frac{1}{\sigma^2} + \theta \alpha \right)^{-1}} \int \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma^2} + \theta \alpha \right) \left( \xi + \frac{\theta \alpha (s - x^B(z, p))}{\frac{1}{\sigma^2} + \theta \alpha} \right)^2 \right\} d\xi \]

\[ = \frac{1}{\sqrt{1 + \theta \alpha \sigma^2}} \exp \left( \frac{1}{2} \frac{\theta \alpha^2}{\sigma^2} + \frac{1}{\sigma^2} + \theta \alpha (s - x^B(z, p))^2 \right) \]

Combining these results, the objective function related to \( p \) is

\[ (s - x^B(z, p)) E_D d + px^B(z, p) - \left( \frac{\theta}{2} V_D d + \frac{1}{2} \frac{\alpha}{1 + \theta \alpha \sigma^2} \right) (s - x^B(z, p))^2. \]

Comparing with Eq. (3.14) completes the proof.

\[ \frac{\theta \alpha}{2} V_D d = \frac{\theta}{2} V_D d + \frac{1}{2} \frac{\alpha}{1 + \theta \alpha \sigma^2}. \]

## C.3 Proof of proposition 3.2

Note that the explicit form of \( x(z, s) = x^B(z, p(z, s)) \) is

\[ x(z, s) = \alpha + \beta A + (\beta B + \gamma) z + \beta BC s. \]

For the first result \( \lim_{\theta \alpha \to \infty} x(z, s) = s \), recall the system of characterizing Eq. \([A]\), \([B]\),
and $[C]$. By taking $\theta_D \to \infty$,

$$[A] : A = -\frac{\alpha}{\beta}, \quad [B] : B = -\frac{\gamma}{\beta}, \quad [C] : C = -\frac{1}{\gamma}.$$ 

In other words, $\beta BC = 1, \beta B + \gamma = 0$, and $\alpha + \beta A = 0$. Hence, as $\theta_D \to \infty$, $x(z,s) \to s$.

For the second result $\lim_{\theta_D \to 0} x(z,s) = 0$, note that from $[C]$ and $\beta I < \beta < 0$, $C \to 0$ as $\theta_D \to 0$. By substituting $C = 0$ into $\{[\alpha], [\beta], [A], [B]\}$,

$$[\alpha] : \alpha = \frac{\kappa_d \bar{d}}{\theta} - (1 - \lambda) \frac{\kappa_e}{\theta} A$$

$$[\beta] : \beta = -\frac{\kappa_d + \kappa_e}{\theta} + (1 - \lambda) \frac{\kappa_e}{\theta} \frac{1}{B}$$

$$[A] : A = -\frac{\alpha}{2\beta} + \frac{\kappa_d \bar{d}}{2(\kappa_d + \kappa_e)}$$

$$[B] : B = -\frac{\gamma}{2\beta} + \frac{\kappa_e}{2(\kappa_d + \kappa_e)}.$$ 

One can check

$$B = \frac{\kappa_e}{\kappa_d + \kappa_e}, \quad \beta = -\lambda \frac{\kappa_d + \kappa_e}{\theta}, \quad \alpha = \lambda \frac{\kappa_d \bar{d}}{\theta}, \quad A = \frac{\kappa_d \bar{d}}{\kappa_d + \kappa_e}$$

solve the system and $\alpha + \beta A = \beta B + \gamma = \beta BC = 0$. Hence, $x(z,s) \to 0$ as $\theta_D \to 0$.

### C.4 Proof of theorem 3.2

We prove each of the four claims in order.

**Price informativeness**
By theorem 3.1 and $[C]$, $C$ is negative and satisfies

$$\frac{1}{C} = \frac{\gamma (\kappa_d + \kappa_\epsilon) - \beta (\theta_D \gamma + \kappa_\epsilon)}{\theta_D \beta} = \frac{\gamma (\kappa_d + \kappa_\epsilon)}{\theta_D \beta} - \frac{\kappa_\epsilon}{\theta_D}.$$  

To show $Q$ decreases, it suffices to show $\beta < 0$ is decreasing in $\theta_D$. The range of $Q$ comes from the range of $C$, $\left(-\frac{1}{\gamma}, 0\right)$.

We show $\beta$ is decreasing by the implicit function theorem. Denote $b(\beta)$ in Eq. (C.6) by $b(\beta, \theta_D)$. The goal is to show

$$\frac{\partial \beta}{\partial \theta_D} = -\frac{\partial b(\beta, \theta_D)/\partial \theta_D}{\partial b(\beta, \theta_D)/\partial \beta} < 0.$$  

Since $b(\beta)$ crosses horizontal line from below at the equilibrium $\beta < 0$, $\frac{\partial b(\beta, \theta_D)/\partial \beta}$ evaluated at the equilibrium $\beta < 0$ is always strictly positive. Hence, by the implicit function theorem, it suffices to check $\frac{\partial b}{\partial \theta_D}$ evaluated at the equilibrium $\beta < 0$ is positive. By using $\beta^3 = -\frac{1}{b_3} (b_0 + b_1 \beta + b_2 \beta^2)$,

$$\frac{\partial b}{\partial \theta_D} = \frac{\partial b_0}{\partial \theta_D} + \frac{\partial b_1}{\partial \theta_D} \beta + \frac{\partial b_2}{\partial \theta_D} \beta^2 + \frac{\partial b_3}{\partial \theta_D} \beta^3 = \frac{\partial b_0}{\partial \theta_D} - \frac{\partial b_3}{\partial \theta_D} b_3 + \left( \frac{\partial b_1}{\partial \theta_D} - \frac{\partial b_3}{\partial \theta_D} b_1 \right) \beta + \left( \frac{\partial b_2}{\partial \theta_D} - \frac{\partial b_3}{\partial \theta_D} b_2 \right) \beta^2.$$  

By direct calculation, with $\Psi := \theta^2 \theta_D^2 + \kappa_\epsilon (\theta^2 + (1 + \lambda) \theta \theta_D + \lambda \theta_D^2) > 0$,

$$\frac{\partial b_0}{\partial \theta_D} - \frac{\partial b_3}{\partial \theta_D} b_3 = -\frac{\kappa_\epsilon \lambda^2 (\kappa_d + \kappa_\epsilon)^3 [2 \theta^2 \theta_D + \kappa_\epsilon \kappa_s \{(1 + \lambda) \theta + 2 \lambda \theta_D\}]}{\theta^2 \Psi} < 0,$$  

178
\[
\frac{\partial b_2}{\partial \theta_D} - \frac{\partial b_3}{\partial \theta_D} \frac{b_2}{b_3} = \frac{\kappa_s^2 \kappa_e}{\theta \Psi} \left[ \kappa_s \kappa_e (\kappa_d + \kappa_e) \left\{ \lambda^2 (\theta + \theta_D)^2 + (\lambda \theta_D + \theta)^2 \right\} + 2 \kappa_e \theta_D \left\{ (1 + \lambda) \theta^2 + \lambda \theta_D \right\} + \kappa_d \theta_D^2 \left\{ 1 + \lambda (2 - \lambda) \right\} + 4 \kappa_d \theta_D^3 \right] > 0,
\]

\[
\frac{\partial b(\lambda \beta_I, \theta_D)}{\partial \theta_D} = \frac{\kappa_d \kappa_s^2 \kappa_e \lambda^2 \theta_D^2 (\kappa_d + \kappa_e)^2 (1 - \lambda) \left\{ 2 \theta + (1 + \lambda) \theta_D \right\}}{\theta \Psi} > 0.
\]

These results imply that the quadratic function \( \frac{\partial b(\beta)}{\partial \theta_D} \) is decreasing and positive on \([\beta_I, \lambda \beta_I]\).

Since the equilibrium \( \beta \) is in this region by theorem 3.1, the proof completes.

**Price volatility**

By direct calculation,

\[
\lim_{\theta_D \to \infty} V(p(z,s)) - \lim_{\theta_D \to 0} V(p(z,s)) = \left( \lim_{\theta_D \to \infty} B^2 - \lim_{\theta_D \to 0} B^2 \right) \left( \kappa_d^{-1} + \kappa_e^{-1} \right) + \left( \lim_{\theta_D \to \infty} (BC)^2 - \lim_{\theta_D \to 0} (BC)^2 \right) \kappa_s^{-1} \kappa_s (\kappa_d + \kappa_e) \left\{ \kappa_s \kappa_e \lambda^2 (\kappa_d + \kappa_e) + (\kappa_d + \lambda \kappa_e) \theta_D \right\}^2
\]

\[
= \theta^2 \kappa_s^2 \lambda^2 (2 \lambda - 1) (\kappa_d + \kappa_e) \kappa_s^2 + \theta^4 \kappa_s \left\{ (2 \kappa_d + \kappa_d) \lambda^2 + \kappa_d (2 \lambda - 1) \right\} \kappa_s + \theta^6 (\kappa_d + \kappa_e)
\]

Note that the denominators of the last expression is positive. From the second equality, if \( 2 \lambda - 1 \geq 0 \), there is no restriction on \( \kappa_s \). If \( 2 \lambda - 1 < 0 \), the numerator of the ratio after the last equality is a concave quadratic function of \( \kappa_s \) with a positive intercept. Hence, by taking small enough \( \kappa_s \), the numerator is positive.

**Welfare of the traders**

Fix a positive \( \theta_D = \bar{\theta}_D > 0 \) and its associated equilibrium price \( p(z,s) \). Note that the traders can always choose demand functions to be identically 0 by selecting \( \alpha_I = \beta_I = \gamma_I = \alpha_O = \)
\[ \beta_O = 0. \text{ Hence, the equilibrium interim utilities satisfy} \]
\[
E \left[ -e^{-\theta(d-p(z,s))x_B^f(z,p(z,s))} | p(z,s) , z \right] \geq E \left[ -e^{-\theta(d-p(z,s))x_B^0 | p(z,s) , z \right] = 1
\]
\[
E \left[ -e^{-\theta(d-p(z,s))x_B^0 | p(z,s) \right] \geq E \left[ -e^{-\theta(d-p(z,s))x_B^0 | p(z,s) = -1. \right]
\]

The right sides of the inequalities are the ex-ante welfare when \( \theta_D \to 0 \) by proposition 3.2.

By taking expectation on both sides and using the tower property, the proof completes.

**Welfare of the dealer**

When \( Y \) is an \( n \times 1 \) vector distributed \( N (\mu, \Sigma) \) and \( A \) is an \( n \times n \) matrix,

\[
EY'AY = E \text{tr} (Y'AY) = tr (AEYY') = tr \{ A (\Sigma + \mu \mu') \}.
\]

Since the profit of the dealer is a quadratic form of uncertainty \( X \), by using this property, we can calculate the expected profit of the dealer in terms of coefficients \( (\alpha, \beta, \gamma, A, B, C) \).

By substituting coefficient values that correspond to \( \theta_D \to \infty \) and \( \theta_D \to 0 \),

\[
\lim_{\theta_D \to 0} u_D - \lim_{\theta_D \to \infty} u_D = \frac{\theta \left( s^2 \kappa_s (\kappa_c \kappa_s \lambda^2 + \theta^2) + \kappa_c \kappa_s \lambda + \theta^2 \right)}{\kappa_s (\kappa_d + \kappa_e) \{ \kappa_s \kappa_e \lambda + \theta^2 (1 + \lambda) \}} > 0.
\]

**C.5 Proof of proposition 3.3**

The proof is by guess and verify, and is composed of three parts: the trader’s problem, the dealer’s problem, and coefficient matching. Guess the equilibrium price is linear

\[
p_t = p(z_t, s_t, x_{t-1}^D) = A_0 + A_x x_{t-1}^D + B (z_t + C's_t).
\]
Note that $x_{t-1}^D$ is the endogenous state.

**Traders’ problem**

Guess the form of value function of the informed traders is, for some nonzero constant $\alpha_I \in \mathbb{R}$ and symmetric matrix $Q_I$,

$$J \left( x_{t-1}, m^I_t; z_t, s_t \right) = -e^{-\alpha_I(m^I_t + p_t x_{t-1}^I) - q_I(z_t, s_t, x_{t-1}^D)}, \quad q_I \left( z_t, s_t, x_{t-1}^D \right) = \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_{t-1}^D \end{bmatrix}' Q_I \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_{t-1}^D \end{bmatrix}$$

where the bold letter $x_{t-1} = (x^I_{t-1}, x^D_{t-1})$ and $p_t$ are functions of $(z_t, s_t)$ and $q_I$ is a quadratic form. Note that although informed traders cannot observe $s^D_t$, $J$ depends on $s^D_t$ through $p_t$.

To be more precise, $J$ depends on all past shocks $(s^D_{\tau})_{\tau \leq t}$ through equilibrium $(p_{\tau}, x^D_{\tau-1})_{\tau \leq t}$. The equilibrium $x^D_t$ is measurable with respect to $\mathcal{F}^I_t$ due to market clearing $x^I_t + x^D_t = \bar{x}$ in equilibrium, so is a constant for informed agents. Given $(x_{t-1}^I, x^D_t, s^I_t, p_t)$, informed trader solves

$$\max_{x_t, y_t, \pi_t} -e^{-\theta_t \pi_t - \beta E_t^I \left[ e^{-\alpha_I(m^{I+1}_t + p_{t+1} x_t) - q_I(z_{t+1}, s_{t+1}, x_{t+1}^D)} \right]} \text{ s.t. } \begin{cases} m^I_{t+1} = d_{t+1} x_t + R y_t + s^I_t d^I_{t+1} \\ \pi_t + y_t + p_t x_t = p_t x^I_{t-1} + m^I_t \end{cases}$$
Since there are three control variables, we solve the problem in three steps. The first step is to substitute out risk-free bond \( y_t \). Note that \( m_{t+1}^I \) can be written as

\[
m_{t+1}^I = (d_{t+1} - R p_t) x_t + R p_t x_{t-1}^I + s_t^I d_{t+1}^I + R m_t^I - R c_t
\]

so that the second term of the objective function can be written as a quadratic function of shocks

\[
E_t^I \left[ e^{-\alpha_t \left\{ (p_{t+1} + d_{t+1} - R p_t) x_t + R p_t x_{t-1}^I + s_t^I d_{t+1}^I + R m_t^I - R c_t \right\} - q_I (z_{t+1}, s_{t+1}, x_t^D)} \right] = E_t^I \exp \left( \begin{bmatrix} 1 \\ d_{t+1} \\ d_t^I \\ z_{t+1} \\ s_{t+1} \end{bmatrix} ^\prime \begin{bmatrix} n_{t0} & \frac{1}{2} n_t^I \\ \frac{1}{2} n_I & \frac{1}{2} N_I \\ \frac{1}{2} n_t & \frac{1}{2} N_t \end{bmatrix} \begin{bmatrix} 1 \\ d_{t+1} \\ d_t^I \\ z_{t+1} \\ s_{t+1} \end{bmatrix} \right) \]

where \( N_I \) is a matrix function of \( Q_I \) and \((n_{t0}, n_I)\) are linear functions of other non-random variables such as \( x_t \). Second, given the normality assumption, this term is an increasing transformation of a quadratic function of \( x_t \), so \( x_t \) can be maximized out

\[
\min_{x_t} n_{t0} + n_t^I \mu_I (z_{t}) + \frac{1}{2} \mu_I (z_{t}) ^\prime N_I \mu_I (z_{t}) + \frac{1}{2} (n_I + N_I \mu_I (z_{t})) ^\prime (\Sigma_I ^{-1} - N_I )^{-1} (n_I + N_I \mu_I (z_{t}))
\]

\[
= - \alpha_I R (m_t^I + p_t x_{t-1}^I - c_t) + \tilde{q}_t^I
\]

182
where \((\mu_I, \Sigma_I)\) are conditional moments

\[
\mu_I(z_t) = E \left[ (d_{t+1}, d_{t+1}', z_{t+1}, s_{t+1})' | z_t \right], \quad \Sigma_I = V \left[ (d_{t+1}, d_{t+1}', z_{t+1}, s_{t+1}) | z_t \right]
\]

and \(\tilde{q}_t^I\) is a quadratic form of \([1, p_t, x_t^D, z_t, s_t^I]\) independent of \(\pi_t\). This step also yields the asset demand

\[
x_t^I = q_{xp} p_t + q_{xz} z_t + q_{xs} s_t^I + q_{x0}
\]

where \((q_{xp}, q_{xz}, q_{xs}, q_{x0})\) are functions of \((A_0, A_x, B, C, Q_I)\). The third step is to maximize over profit \(\pi_t\) the objective function

\[
\max_{\pi_t} -e^{-\theta_I \pi_t} - \beta |I - \Sigma_I N_t|^{-\frac{1}{2}} e^{-\alpha_I R (m_I + p_t x_{t-1}^I - \pi_t) + \tilde{q}_t^I}
\]

By taking first order condition with respect to \(\pi_t\),

\[
\pi_t = \frac{1}{\theta_I + \alpha_I R} \ln \left( \frac{\theta_I}{\beta \alpha_I R |I - \Sigma_I N_t|^{-\frac{1}{2}}} \right) + \frac{\alpha_I R}{\theta_I + \alpha_I R} (m_I + p_t x_{t-1}^I) - \frac{1}{\theta_I + \alpha_I R} \tilde{q}_t^I.
\]

Substituting \(\pi_t\) into the objective function leads to

\[
J(x_{t-1}, m_I^t, z_t, s_t)
\]

\[
= - \beta |I - \Sigma_I N_t|^{-\frac{1}{2}} \left( \frac{\alpha_I R + \theta_I}{\theta_I} \right) \left( \frac{\theta_I}{\beta \alpha_I R |I - \Sigma_I N_t|^{-\frac{1}{2}}} \right)^{\frac{\alpha_I R}{\theta_I + \alpha_I R}}
\]

\[
\times \exp \left( - \frac{\theta_I \alpha_I R}{\theta_I + \alpha_I R} (m_I + p_t x_{t-1}^I) + \frac{\theta_I}{\theta_I + \alpha_I R} \tilde{q}_t^I \right)
\]
This has to be identical to the original guess so that

\[
\frac{\theta_I \alpha_I R}{\theta_I + \alpha_I R} = \alpha_I \Rightarrow \alpha_I = \frac{R - 1}{R} \theta_I.
\]

To get the equation that characterizes \(Q_I\), note that \(\tilde{q}_t^I = \tilde{q}^I (p_t, x_t^D, z_t, s_t^I)\) can be written as a quadratic form with respect to the same variables as \(q^I (z_t, s_t, x_t^{D-1})\).

\[
\begin{cases}
x_t^I = q_p p_t + q_{xz} z_t + q_{x} s_t^I + q_{x0} \\
x_t^I + x_t^D = \bar{x} \\
p_t = A_0 + A_x x_{t-1}^D + B (z_t + C' s_t)
\end{cases} \Rightarrow \tilde{q}^I (p_t, x_t^D, z_t, s_t^I) = \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_{t-1}^D \end{bmatrix}^T \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_{t-1}^D \end{bmatrix}
\]

Substituting \(\alpha_I\) into the objective function and comparing it with the original guess lead to

\[
Q_I = - \left[ \frac{1}{R} \tilde{Q}_I + \left\{ \frac{1}{R} \ln \left( |I - \Sigma_I N_I|^{-\frac{1}{2}} \beta (R - 1) \right) + \ln \frac{R}{R - 1} \right\} i_{11} \right]
\]

where \(i_{11}\) is \(5 \times 5\) matrix with \((1, 1)\) element being 1 and all other elements being 0. Since \(\tilde{Q}_I\) is a function of \(Q_I\), this equation pins down the fixed point.
Dealer’s problem

Guess the value function of the dealer to be, for some nonzero constant $\alpha_D \in \mathbb{R}$ and symmetric matrix $Q_D$,

$$J \left( x_{t-1}^D, m_t^D, s_t^D, x_t^I \left( z_t, s_t^I, \cdot \right) \right) = -e^{-\alpha_D m_t^D - q_D \left( z_t, s_t^I, x_t^I \right)}$$

where $q_D \left( z_t, s_t^I, x_t^I \right) = \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_t^D \\ x_t^{D-1} \end{bmatrix}' Q_D \begin{bmatrix} 1 \\ z_t \\ s_t \\ x_t^D \\ x_t^{D-1} \end{bmatrix}$

Dealers do not observe $s_t^I$, but it affects value through the demand schedule

$$x_t^I = q_{xp} p_t + q_{xz} \hat{z}_t + q_{xs} s_t^I + q_x 0 = q_{xp} p_t + q_{xz} \left( z_t + q_{xs} s_t^I \right) + q_x 0$$

where $\hat{z}_t$ is observable to the dealer. The dealer’s problem is then

$$\max_{x_t, y_t, \pi_t, p_t} -e^{-\theta_D \pi_t} - \beta E^D \left[ e^{-\alpha_D m_{t+1}^D - q_D \left( z_{t+1}, s_{t+1}^I, x_{t+1}^I \right)} \right] \text{s.t.} \begin{cases} m_{t+1}^D = d_{t+1} x_t + R y_t + s_t^D d_{t+1} \\ \pi_t + y_t + p_t x_t = p_t x_{t-1}^D + m_t^D \\ x_t = \bar{x} - \left( q_{xp} p_t + q_{xz} \hat{z}_t + q_x 0 \right) \end{cases}$$

The steps to solve the problem is similar to those in the traders’ problem, except that the choice variable contains price. First, by deleting risk-free bond,

$$m_{t+1}^D = (d_{t+1} - R p_t) x_t + R p_t x_{t-1}^D + s_t^D d_{t+1} + R m_t^D - R c_t$$
The second term of the objective function becomes

\[
E_t^D e^{-\alpha D \left\{ (d_{t+1} - R_p) x_t + R_p x_{t-1} P + s_t P d_{t+1} + R m_t - R c_t \right\} - q_D (z_{t+1}, s_{t+1}, x_t)}
\]

\[
= E_t^D \exp \left( \begin{bmatrix} 1 \\ d_{t+1} \\ d_{t+1}^P \\ z_{t+1} \\ s_{t+1} \end{bmatrix} \right)' \begin{bmatrix} n_{D0} \\ \frac{1}{2} n_D' \\ \frac{1}{2} n_D \\ \frac{1}{2} N_D \end{bmatrix} \left( \begin{bmatrix} 1 \\ d_{t+1} \\ d_{t+1}^P \\ z_{t+1} \\ s_{t+1} \end{bmatrix} \right) \right)
\]

\[
\propto n_{D0} + n_{D}' \mu_D (\hat{z}_t) + \frac{1}{2} \mu_D (\hat{z}_t)' N_D \mu_D (\hat{z}_t) + \frac{1}{2} (n_D + N_D \mu_D (\hat{z}_t))' (\Sigma_D^{-1} - N_D)^{-1} (n_D + N_D \mu_D (\hat{z}_t)),
\]

where \( N_D \) is a matrix function of \( Q_D \), \( n_{D0} \) is a linear function of \((x_t, x_{t-1}^2, p_t x_t)\), \( n_D \) is a linear function of \( x_t \),

\[
\mu_D (\hat{z}_t) = E \left[ [d_{t+1}, d_{t+1}^P, z_{t+1}, s_{t+1}]' | \hat{z}_t \right], \quad \Sigma_D = V \left[ [d_{t+1}, d_{t+1}^P, z_{t+1}, s_{t+1}] | \hat{z}_t \right] .
\]

By substituting the demand schedule, the problem reduces to a minimization of a quadratic function of \( p_t \), which is the second step. Hence, the second term of the objective function an
increasing function of

\[
\min_{p_t} n_{D0} + n_D' \mu (\hat{z}_t) + \frac{1}{2} \mu (\hat{z}_t)' N_D \mu (\hat{z}_t) + \frac{1}{2} (n_D + N_D \mu (\hat{z}_t))' \left( \Sigma_D^{-1} - N_D \right)^{-1} (n_D + N_D \mu (\hat{z}_t))
\]

\[
= - \alpha_D R (m^D_t - c_t) + \begin{bmatrix}
1 \\
\hat{z}_t \\
\hat{s}_t \\
x^D_{t-1}
\end{bmatrix} \begin{bmatrix}
1 \\
\hat{z}_t \\
\hat{s}_t \\
x^D_{t-1}
\end{bmatrix} = \tilde{Q}_D
\]

and the optimal pricing is

\[
p_t = q_{p0} + q_{pz'} \hat{z}_t + q_{ps'} \hat{s}_t + q_{px'} x^D_{t-1}
\]

and \((q_{p0}, q_{pz}, q_{ps}, q_{px})\) are functions of \((q_{xp}, q_{xz}, q_{xs}, q_{x0}, Q_D)\). The third step is to choose \(\pi_t\) to solve

\[
\max_{\pi_t} -e^{-\theta_D \pi_t} - \beta |I - \Sigma_D N_D|^{-\frac{1}{2}} e^{-\alpha_D R (m^D_t - \pi_t) + \tilde{q}_D^P}.
\]

By taking the first order condition,

\[
\pi_t = \frac{1}{\theta_D + \alpha_D R} \ln \left( \frac{\theta_D}{\beta \alpha_D R |I - \Sigma_D N_D|^{-\frac{1}{2}}} \right) + \frac{\alpha_D R}{\theta_D + \alpha_D R} m^D_t - \frac{1}{\theta_D + \alpha_D R} \tilde{q}_D^P.
\]

Since the implied value function has to coincide with the original guess,

\[
\alpha_D = \frac{R - 1}{R} \theta_D, \quad Q_D = - \left[ \frac{1}{R} \tilde{Q}_D + \left\{ \frac{1}{R} \ln \left( |I - \Sigma_D N_D|^{-\frac{1}{2}} \beta (R - 1) \right) + \ln \frac{R}{R - 1} \right\} i_{11} \right].
\]
Once \((\alpha_D, Q_D)\) is obtained, the profit becomes

\[
\pi_t^D = \frac{-1}{R\theta_D} \ln \left( \beta (R - 1) |I - \Sigma_D N_D|^{-\frac{1}{2}} \right) + \frac{R - 1}{R} m_t^D - \frac{1}{R\theta_D} q_t^D
\]

and the coefficients of the pricing decision becomes a function of \((q_{xp}, q_{xz}, q_{xs}, q_{x0})\) and has to coincide with the initial guess

\[
p_t = q_{p0} + q_{pz} \tilde{z}_t + q_{ps} s_t^D + q_{px} x_{t-1}^D = A_0 + A_x x_{t-1}^D + B (z_t + C's_t).
\]

**Fixed point**

Given the price coefficients \((A_0, A_x, B, C)\), the informed agents’ problem gives a fixed point equation of \(Q_I\). \((A_0, A_x, B, C, Q_I)\) determines coefficients of \(x^I\), \((q_{xp}, q_{xz}, q_{xs}, q_{x0})\). Given \((q_{xp}, q_{xz}, q_{xs}, q_{x0})\), the dealer’s problem gives a fixed point equation for \(Q_D\). \((q_{xp}, q_{xz}, q_{xs}, q_{x0}, Q_D)\) then determines \((A_0, A_x, B, C)\). One can solve the fixed point \((Q_I, Q_D, A_0, A_x, B, C, q_{xp}, q_{xz}, q_{xs}, q_{x0})\) numerically by repeating the loop until convergence. Once the fixed point is obtained, the transition of \(x_t^D\) can be obtained from

\[
p_t = A_0 + A_x x_{t-1}^D + B (z_t + C's_t), \quad x_t^I = q_{xp} p_t + q_{xz} z_t + q_{xs} s_t^I + q_{x0}, \quad x_t^D = \bar{x} - x_t^I.
\]

By substituting \(x_t^I\) and \(p_t\) out, the transition of \(x_t^D\) is obtained by

\[
x_t^D = \bar{x} - q_{x0} - q_{xp} A_0 - (q_{xp} B + q_{xz}) \bar{d} - q_{xp} A_x x_{t-1}^D - (q_{xp} B + q_{xz}) (z_t - \bar{d}) - q_{xp} B C's_t - q_{xs} s_t^I.
\]
One can define

$$\rho_0 := \bar{x} - q_{x0} - q_{xp} A_0 - (q_{xp} B + q_{xz}) \bar{d}, \quad \rho_1 := -q_{xp} A_x, \quad \epsilon^D_i := - (q_{xp} B + q_{xz}) (z_i - \bar{d}) - q_{xp} BC' s_i - q_{xs} s^I_i.$$