

# ESSAYS IN FINANCIAL ECONOMICS

LIANG DAI

A DISSERTATION  
PRESENTED TO THE FACULTY  
OF PRINCETON UNIVERSITY  
IN CANDIDACY FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE  
BY THE DEPARTMENT OF  
ECONOMICS

ADVISER: PROFESSOR STEPHEN E. MORRIS

JUNE 2015

© Copyright by Liang Dai, 2015.

All rights reserved.

# Abstract

This dissertation studies three different topics in financial economics.

The first chapter investigates how a profit-maximizing asset originator can coordinate the information acquisition of and interaction among investors with different expertise by means of asset bundling. Bundling is beneficial to the originator when it discourages investors from analyzing idiosyncratic risks and focuses their attention on aggregate risks. But it is optimal to sell aggregate risks separately in order to exploit investors' heterogeneous expertise in learning about them and thus lower the risk premium. This analysis rationalizes the common securitization practice of bundling loans by asset class, which is at odds with existing theories based on diversification. The analysis also offers an alternative perspective on conglomerate formation (a form of asset bundling), and its relation to empirical evidence in that context is discussed.

The second chapter, coauthored with Cheng Chen, proposes and provides empirical support for a novel explanation for the middle income trap: political distortion intensifies with economic development and impedes further growth. A parsimonious model is built to highlight the tradeoff faced by the policy maker between public interest of economic development and private interest of special interest groups. The latter dominates the former when the economy is good, as special interest groups benefit disproportionately from economic development and tilt policy making towards them by providing greater political rents. We use empirical evidence from the historical episode of the removal of branching restriction of U.S. banking industry to support our hypothesis.

The third and last chapter studies the contract design implication of routine auditing. In real life, usually people routinely pay accounting firms upfront before the publication of the financial reports of a firm being audited. I show in a model that otherwise resembles Gale and Hellwig (1985) that the optimal contract is equity instead of debt. This contrasts many existing security design models and points out state contingency of auditing technology is the driving force of their result.

## Acknowledgements

I would like to express my deepest gratitude to my adviser Stephen Morris for his guidance, patience, support and invaluable insights. It has been an honor and a privilege to be his student, and a wonderful experience to learn from him.

I would also like to express my sincere gratitude to Hyun Song Shin, Wei Xiong and Valentin Haddad for their tremendous support, thoughtful advice and wisdom. It has been a great pleasure to learn from them, and I am very grateful for all their effort.

Thank you to my classmates and friends Qingqing Cao, Cheng Chen, Shaowei Ke, Xiaochen Feng, Ji Huang, Zongbo Huang, Jason Ravit Chang Sun, Thomas Winberry, Christian Vom Lehn, Ming Yang and Yu Zhang, who made my study at Fisher Hall infinitely more pleasant.

I am also indebted to professors and fellow students who attended my presentations at finance, microeconomic theory, or political economy lunch seminars and gave valuable comments on my work. Special thanks are given to Sylvain Chassang, Roland Benabou, Markus Brunnermeier, Maryam Farboodi, Nobuhiro Kiyotaki, Oleg Itskhoki, Benjamin Moll, Kristopher Ramsey and Richard Rogerson.

Thank you to all my non-econ friends, for making Princeton like home.

Finally, thank you to my parents and grandparents, to whom I owe everything.

To my parents.

# Contents

Abstract . . . . .	iii
Acknowledgements . . . . .	v
List of Tables . . . . .	viii
<b>1 Asset Bundling and Information Acquisition of Investors with Dif-</b>	
<b>ferent Expertise</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Baseline Model . . . . .	7
1.2.1 Factors and Asset Payoffs . . . . .	7
1.2.2 Originator . . . . .	7
1.2.3 Investors . . . . .	8
1.2.4 Noise Trader . . . . .	13
1.2.5 Equilibrium . . . . .	14
1.2.6 Summary of Model Setup . . . . .	14
1.3 The Upside of Bundling: The Discipline Channel . . . . .	16
1.4 The Downside of Bundling: The Trade-Restriction Channel and the Specialization-Destruction Channel . . . . .	21

1.5	Discussion . . . . .	26
1.5.1	A Generalization: The Optimality of Categorization Strategy . . . . .	26
1.5.2	Relation to the Security-Design Literature . . . . .	31
1.5.3	Optimal Bundling Strategies May Not Favor Investors . . . . .	32
1.5.4	Extrapolation Between the Two Polar Cases . . . . .	34
1.6	Asset Bundling and Corporate Diversification . . . . .	35
1.6.1	A Market-Side Theory of Corporate Diversification . . . . .	35
1.6.2	Empirical Predictions . . . . .	38
1.7	Conclusion . . . . .	40
1.8	References . . . . .	41
1.9	Appendix . . . . .	45
1.9.1	Problems and Propositions in the n-factor setup . . . . .	45
1.9.2	Propositions in the 2-factor setup . . . . .	58
<b>2</b>	<b>The Middle Income Trap, Branching Deregulation, and Political Influence</b>	<b>66</b>
2.1	Introduction . . . . .	66
2.2	The Model . . . . .	74
2.2.1	Environment . . . . .	75
2.2.2	Financial Friction and Credit Rationing . . . . .	78
2.2.3	Political Contribution and Branching Deregulation . . . . .	81
2.2.4	Implications for the Middle Income Trap . . . . .	85
2.3	Empirical Evidence . . . . .	87



2.3.1	Deregulation and GDP . . . . .	87
2.3.2	Banks' Profit and GDP . . . . .	87
2.3.3	Deregulation, Net Interest Margin and Banks' Profit . . . . .	88
2.3.4	Banks' Profit and Campaign Contribution . . . . .	91
2.4	Conclusion . . . . .	94
2.5	References . . . . .	96
<b>3</b>	<b>Contract Design under Routing Auditing</b>	<b>99</b>
3.1	Introduction . . . . .	99
3.2	The Model . . . . .	103
3.3	Equity as Optimal Contract . . . . .	106
3.4	Conclusion . . . . .	111
3.5	References . . . . .	111

# List of Tables

- 2.1 Growth Rate of Bank Profit and GDP at the State-Level . . . . . 88
- 2.2 Deregulation and Profit in the Banking Industry . . . . . 92
- 2.3 Political Contributions and Bank Profit . . . . . 94

# Chapter 1

## Asset Bundling and Information Acquisition of Investors with Different Expertise

### 1.1 Introduction

Securitization plays an important role in the U.S. economy. As of April 2011, outstanding securitized assets totalled \$11 trillion, which was substantially more than the amount of all outstanding marketable U.S. Treasury securities (Gorton and Metrick, 2013). One salient feature of securitization is that the creation of asset-backed securities (ABS) always involves pooling loans of the same asset class; i.e., a pool consists exclusively of mortgages, auto receivables or credit card receivables. Different asset classes are not mixed, even if the originator in fact instigates loans of many

different asset classes. Existing theories based on diversification<sup>1</sup> do not square well with this feature, as one would expect the benefit of diversification to be greater when different asset classes are mixed.

In this paper I demonstrate that this feature is no longer a puzzle if we recognize the important role played by the heterogeneous expertise of investors in acquiring different information about asset payoffs, which existing theories of securitization abstract from. Pooling all loans of the same asset class prohibits buyers from cherry-picking individual loans, and thus prevents them from using their expertise to exploit other buyers regarding the risk peculiar to the loans picked. This encourages all buyers to acquire information only about risks common to all the loans being sold. Since they face less uncertainty after learning about these risks, buyers demand a lower risk premium from the originator. Different asset classes are sold separately. This enables mortgage specialists to freely trade mortgages and to profit from mortgage-specific information and thus induces them to specialize in acquiring information in their area of expertise. The comparative advantages in information acquisition of different buyers are thus better utilized and fewer risks are priced, benefiting the originator.

This paper develops a model that formalizes this explanation, and further studies a broader theoretical issue: How can a self-interested asset originator coordinate the

---

<sup>1</sup>For example, Subrahmanyam (1991) shows that the introduction of a basket of securities reduces the problem of adverse selection by offsetting demand from informed traders that have private information about individual securities. Demarzo (2005) shows that when the seller has better information, pooling makes the value of the ABS created less sensitive to the his private information about individual assets. When instead the buyer has better information, pooling prevents her from cherry-picking only good assets. Adverse selection is reduced in both cases, as private information about individual assets is diversified.

information acquisition of and interaction among investors that have *different areas of expertise*? Because potential investors in any financial asset inherently have different learning expertise, this seems to be a fundamental question in understanding the workings of financial market, in addition to rationalizing the puzzle as an application, but has received little attention in the literature to date. As a first step, this paper focuses on asset bundling, a technique commonly used by asset originators. The application of asset bundling in financial market practice is not limited to securitization. Indeed, a conglomerate can also be viewed as a bundle of its several lines of businesses, in the sense that its stakeholders cannot selectively invest in and receive cash flows from any particular business that it operates. Thus, the model developed can also be used to study conglomerate formation.

My model features two key ingredients: the interaction of heterogeneous investors and their endogenous learning behavior. Asset payoffs are determined by different risks; e.g., sector-specific risk, region-specific risk, asset-specific risk. There is one asset originator and a continuum of investors with different learning expertise. Each risk-averse investor allocates his limited attention to learning about these risks before trading the assets. How he does that is endogenously shaped by the bundling choice of the asset originator and by his interaction with other investors. The asset originator, who wants to maximize the revenue of the sale, bundles his original assets to channel the allocation of investors' learning capacity in the way that minimizes the total risk premium.

Three key theoretical channels novel in the literature are highlighted in the model, leading to the upside and downside of asset bundling.

The upside of asset bundling is driven by a discipline channel: asset bundling restricts speculation on risks that are supposedly diversified away, and gives investors less incentive to acquire information about them. As such, the originator successfully persuades investors to learn only about risks that cannot be reduced by diversification. Since investors face fewer such risks after studying them, they demand a lower risk premium in equilibrium, benefiting the originator.

The downside of asset bundling is driven by two different economic forces. First, asset bundling mechanically restricts the asset span available to investors, thus preventing them from holding their respective favorite portfolios. Hence in equilibrium, they demand lower prices to compensate. This is a trade-restriction channel. Second, asset bundling induces each investor to specialize less in acquiring information about the risk that he has expertise in. Because the expertise of investors is less utilized, there are more risks priced in equilibrium. This is a specialization-destruction channel.

These theoretical channels work not only in the context of securitization, but also in the context of conglomerate formation. By relabelling the asset originator as an entrepreneur who owns several lines of businesses and decides how to set the firm boundaries, my model can also be viewed as one of corporate diversification. It offers a new investor-side (instead of firm-side) perspective of conglomerate formation that can generate both a diversification premium (by the discipline channel) and a discount (by the trade-restriction channel and the specialization-destruction channel), and yields empirical predictions consistent with existing evidence in the literature. As such, my model also builds a conceptual connection between securitization and

conglomerate formation, two seemingly remote contexts that are both important in their own right.

My model follows Van Nieuwerburgh and Veldkamp (2009, 2010), which study the endogenous information acquisition of investors with heterogeneous expertise, and uses their modelling approach. My work differs from theirs, as my focus is on the implications of asset design and asset pricing rather than on the portfolio choices of individual investors.

There are a few papers that also study the endogenous information acquisition of investors. Peng and Xiong (2006) show how the limited attention of a representative investor leads to categorical learning and return comovement. For technical simplification, they do not incorporate the interaction of heterogeneous investors. Subrahmanyam (1991) demonstrates how markets of baskets of securities reduce adverse selection cost. Recently, Goldstein and Yang (2014) identify strategic complementarities in the trading and information acquisition of investors informed about different components of the same asset. These two papers endow traders with exogenous information in their baseline models, and traders are *ex ante* identical in the extensions with endogenous information acquisition.

My work is also related to the literature on security design. In addition to rationalizing the feature of bundling loans by asset classes of securitization, my model complements this literature in two aspects. First, it studies the interaction of heterogeneous security buyers, which existing security design models (e.g. Demarzo and Duffie 1999; Demarzo 2005) typically abstract from. Second, existing security-design models (e.g. Townsend 1979; Dang, Gorton, and Holmstrom 2013) usually focus on

the extensive margin of information acquisition; i.e., how to reduce the costly information acquisition of security buyers. My model focuses instead on the intensive margin: given the resources available to security buyers for information acquisition, how can the seller induce buyers to use those resources in his preferred way? A more detailed discussion on the relation of my work to this literature is given in Section 5.2.

Lastly, my work complements the literature on corporate diversification by offering an alternative perspective on conglomerate formation. A detailed discussion can be found in Section 6.

The rest of this paper is organized as follows. Section 2 introduces the setup of the baseline model. Section 3 illustrates the discipline channel by studying a polar case in which only one risk factor is non-diversifiable. Section 4 illustrates the trade restriction channel and the specialization destruction channel by studying another polar case in which all risk factors are non-diversifiable and play a symmetric role. Section 5 introduces a generalization of the baseline model that establishes the optimality of categorization strategy, and discusses several issues of the baseline model. Section 6 discusses the application of the model in the context of corporate diversification and relevant empirical evidence in the existing literature. Section 7 concludes.



## 1.2 Baseline Model

### 1.2.1 Factors and Asset Payoffs

There are two orthogonal risk factors:  $f_1, f_2$ , and two risky assets, with a supply of one each, and payoffs  $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$ , or in matrix form,  $\mathbf{X} = \Gamma \mathbf{f}$ , such that  $\Gamma = (\gamma_{ij})$  is an orthogonal matrix<sup>2</sup>.  $w_i \equiv \gamma_{1i} + \gamma_{2i}, i = 1, 2$  is the loading of total asset payoff ( $X_1 + X_2$ ) on  $f_i$ . The orthogonality of  $\Gamma$  implies that  $w_1^2 + w_2^2 = 2$ . Without loss of generality, hereafter we consider only the range in which  $w_1 \geq 1$  and  $0 \leq w_2/w_1 \leq 1$ .

There is also a risk-free asset with an unlimited supply, and its gross return is normalized to 1.

### 1.2.2 Originator

There is a risk-neutral originator, who owns all the risky assets and wants to sell them.<sup>3</sup> His objective is to maximize the expected total revenue. To do so, he chooses how to bundle the assets (i.e., creating new tradable non-redundant asset(s) that are

---

<sup>2</sup>The orthogonality of  $\Gamma$  is without loss of generality, as one can always make this true by redefining factors through the eigenvalue decomposition of  $Var(X)$ .

<sup>3</sup>In this paper, the originator is modeled as a monopoly seller of the risky assets. This is for modeling convenience as a first step in studying how an asset originator coordinates the information acquisition of and interaction among investors with different expertise. In reality, although many banks simultaneously issue loans of the same asset class, they typically have different focuses of business. For instance, Citi's businesses within the U.S. are quite heavily concentrated in the New York area, and it has large presence in Mexico through Banamex, while Wells Fargo is headquartered in San Francisco and is typically thought of as a West Coast bank. Thus, a bank can be approximately viewed as a monopoly in its specialties. It would be interesting to extend this model to an oligopoly or a monopolistic competition setup, which is beyond the scope of this paper.

linear combinations of the original assets, such that the former completely absorb the latter), and then sells all of them.<sup>4</sup> This means he can create a single new asset with payoff  $Y = X_1 + X_2$  and supply of 1, or instead he can create two new assets, each with supply of 1, and payoffs  $Y_k = t_{k,1}X_1 + t_{k,2}X_2$ ,  $k = 1, 2$ , such that  $t_{1,i} + t_{2,i} = 1$ ,  $i = 1, 2$ , i.e. the original assets are exhausted, and  $t_{1,1}/t_{2,1} \neq t_{1,2}/t_{2,2}$ .

Each bundling strategy can be uniquely represented by a matrix  $T$ :  $T = (1, 1)$  if a single asset is created, and  $T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$  if two tradable assets are created. By construction,  $T$  has full rank,  $\mathbf{1}'T = \mathbf{1}'$ , and payoff(s) of the tradable asset(s)  $\mathbf{Y} = T\mathbf{X}$ . The originator's problem can be expressed as  $\max_T E_0[\mathbf{1}'\mathbf{p}_T]$ , where  $\mathbf{p}_T$  denotes the price(s) of asset(s) formed by strategy  $T$ .

### 1.2.3 Investors

There are two types  $i \in \{1, 2\}$  of risk-averse investors, each with a continuum of mass  $1/2$ . Each investor starts with a flat prior with mean zero about the factors  $\mathbf{f}$ , and does two things sequentially after observing the bundling choice of the originator: 1) acquires information about the factors  $\mathbf{f}$  to maximize his expected utility at the trading stage; 2) chooses a portfolio of tradable assets  $\mathbf{q}$  to maximize his mean-variance utility:

$$\max_{\mathbf{q}} -E[-\rho\mathbf{q}'(\mathbf{Y} - \mathbf{p}) + \frac{\rho^2}{2}\mathbf{q}'Var(\mathbf{Y})\mathbf{q}]. \quad (1.1)$$

---

<sup>4</sup>Unlike the security-design literature, here it is assumed that the originator cannot retain any asset. This precludes signaling and focuses on the effect of bundling on the information acquisition choice of the investors.

## Expertise and Information Acquisition

Modeling of investors' information acquisition of is based on Van Nieuwerburgh and Veldkamp (2009). Before choosing a portfolio, each investor observes two unbiased private signals about factors  $\mathbf{f}$ . One signal has exogenous precision, and the investor is to choose the precision of the other. Conditional on  $\mathbf{f}$ , signals are independent across investors.

The exogenous signal models the different expertise of investors. Specifically, investor  $\alpha$  of type  $i$ 's (hereafter  $(\alpha, i)$ ) exogenous signal  $\mathbf{s}^{\alpha,i} \sim N(\mathbf{f}, (\Lambda_0^i)^{-1})$ ,<sup>5</sup> where  $\Lambda_0^i = \text{diag}(\lambda_{0,1}^i, \lambda_{0,2}^i)$ . It is assumed that  $\lambda_{0,i}^i = \bar{\lambda} > \underline{\lambda} = \lambda_{0,-i}^i > 0$ ; i.e., from their exogenous signals, type  $i$  investors know  $f_i$  better than others.

The endogenous signal  $\boldsymbol{\eta}^{\alpha,i} \sim N(\mathbf{f}, (\Lambda_\eta^{\alpha,i})^{-1})$  models investors' information acquisition. To highlight the role of expertise, and also for tractability of belief aggregation, it is assumed that its precision matrix  $\Lambda_\eta^{\alpha,i}$  is diagonal, as in Van Nieuwerburgh & Veldkamp (2009) and (2010). This rules out the possibility that an investor chooses to observe a signal correlated with more than one factor. Thus, investors can choose how much to learn about each factor, but are not allowed to change the risk factor structures.

Choosing the precision  $\Lambda_\eta^{\alpha,i}$  is equivalent to choosing the precision of the posterior after observing both signals,  $\Lambda^{\alpha,i} = \text{diag}(\lambda_1^{\alpha,i}, \lambda_2^{\alpha,i}) \equiv \Lambda_\eta^{\alpha,i} + \Lambda_0^i$ . Each investor  $(\alpha, i)$  faces two constraints in this choice:

---

<sup>5</sup>Hereafter, superscripts index investors and subscripts index objects to learn and trade. two-dimensional superscripts are needed to distinguish investors, as different investors of the same type are allowed to behave differently.

1) A capacity constraint that limits the quantity of information carried by the endogenous signals, measured by Shannon capacity, to be no more than  $K$ ,  $K \geq 1$ :

$$\prod_j \lambda_j^{\alpha,i} \leq K \prod_j \lambda_{0,j}^i. \quad (1.2)$$

2) A no-forgetting constraint that prevents the investor from forgetting previous exogenous information about one factor in order to free up capacity to learn about other factors:

$$\lambda_j^{\alpha,i} \geq \lambda_{0,j}^i \quad \forall j \quad (1.3)$$

Note that when  $K = 1$ , the only possible choice of  $\Lambda^{\alpha,i}$  that satisfies both constraints is  $\lambda_j^{\alpha,i} = \lambda_{0,j}^i \quad \forall j$ , which means investors cannot acquire information.

### Comparative Advantage in Information Acquisition

The capacity constraint is a bound on entropy reduction, an information measure with a long history in information theory (Shannon 1948). It is a common distance measure in econometrics (a log-likelihood ratio) and in statistics (a Kullback-Liebler distance), and is used widely in the recent economics literature on rational inattention (see Sims (2010) for a review).

One property of this technology is that,  $K\lambda - \lambda$ , the gain of signal precision by using a given capacity  $K$ , increases with prior knowledge  $\lambda$ . This turns initial information advantage ( $\bar{\lambda} > \underline{\lambda}$ ) into comparative advantage in acquiring additional information:

---

<sup>6</sup>This comes from  $\det[(\Lambda_0^i)^{-1}] / \det[(\Lambda^{\alpha,i})^{-1}] \leq K$ .

1) For a given investor, the marginal gain of signal precision of a factor from additional input of capacity increases with capacity already used on that factor;

2) For a given factor  $f_i$ , the gain of signal precision of type  $i$  is greater than that of other types from the same input of capacity.

In the real world, there is usually a fixed cost for acquiring information about a new asset class or market, because basic background knowledge, skills and equipments have to be developed or acquired upfront. This is a major reason for the difference of expertise of market participants. Such a fixed cost justifies the aforementioned comparative advantage, which is important for the results of this paper.

### **Portfolio Choice**

Investors trade the assets available as in the markets of Admati (1985). Before portfolio choice, each investor observes the realization of his private signals and market price(s)  $\mathbf{p}$  of the tradable assets, which is an additional public noisy signal of  $\mathbf{Y}$ , the payoff(s) of these assets. In the appendix it is shown that  $(\mathbf{p}_T - \mathbf{A}_T) \sim N(\mathbf{Y}, \Omega_{p,T}^{-1})$ , where  $\mathbf{A}_T$  and  $\Omega_{p,T}$  are deterministic objects that depend on the bundling strategy  $T$ . The investor updates his belief about  $\mathbf{Y}$  using Bayes Law and decides how much of each asset to buy,  $\mathbf{q}^{\alpha,i}$ , to maximize his utility (equation 1.1). The technical details of the pricing formula and of investors' portfolio choice are given in the appendix.

### **The Role of Preference**

The mean-variance preference (equation 1.1) follows from risk aversion at the trading stage and from preference for early resolution of uncertainty at the learning stage.

Specifically, an investor's utility function can be expressed as  $U = E_1[u_1(E_2[u_2(W)])]$ , where  $W = W_0 + \mathbf{q}'(\mathbf{Y} - \mathbf{p})$  denotes terminal wealth, the sum of initial wealth  $W_0$  and profit from portfolio investment.

Time 2 refers to the trading stage.  $u_2(W) = -\exp(-\rho W)$ .  $u_2'' < 0$  governs his risk aversion at the trading stage.

Time 1 refers to the learning stage.  $u_1(x) = -\log(-x)$ . Since  $u_1'' > 0$ , the investor prefers early resolution of uncertainty before trading stage: At the learning stage, the investor anticipates that the additional information he will get later may signal either high or low expected utility  $E_2[u_2(W)]$  he will enjoy at trading stage. Therefore, at the learning stage, he sees  $E_2[u_2(W)]$  as a random variable, and has expected utility  $E_1[u_1(E_2[u_2(W)])]$ . If he instead cannot see the additional information before trading, his expected utility at the learning stage is  $E_1[u_1(u_2(W))]$ . Since  $u_1'' > 0$ , Jensen's inequality implies  $E_1[u_1(E_2[u_2(W)])] > E_1[u_1(u_2(W))]$ , i.e. the investor likes to resolve uncertainty by learning before the trading stage.

The preference for early resolution of uncertainty at the learning stage makes the investor choose to learn more about those factors he expects to hold more at the trading stage. His risk aversion at the trading stage makes him hold more of the factors he knows better. These two preferences form a feedback loop and reinforce each other, pressuring the investor to specialize in learning about a single tradable factor. This is crucial for the result of the paper.

## 1.2.4 Noise Trader

As in a standard rational expectation equilibrium model (e.g., Grossman and Stiglitz 1980), noise traders, who trade assets for non-speculative reasons, such as liquidity needs, are needed to prevent investors from being able to perfectly infer the private information of others from prices and thus have no need to acquire any private information themselves.

In this model, different bundling strategies create different tradable assets and possibly different asset spans, and mechanically these strategies require different noise structures for market clearing. Thus, the choice of bundling strategies can also affect the originator's payoff through its impact on noise structure. To make the results of the model neutral to this uninteresting channel, we need to build a connection between these different noise structures. To do this, a representative noise trader is introduced. Her "true" demand for original assets  $\mathbf{X}$  is  $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ ,<sup>7</sup> and her effective demand  $\boldsymbol{\varepsilon}_T$  for tradable asset(s)  $\mathbf{Y} = T\mathbf{X}$  is assumed to be the linear projection of her desired portfolio  $\boldsymbol{\varepsilon}'\mathbf{X}$  onto the tradable assets span:  $\boldsymbol{\varepsilon}_T = (TT')^{-1}T\boldsymbol{\varepsilon} \sim N(0, \sigma^2(TT')^{-1})$ , which is the closest available substitute. As such, noise structures following different bundling strategies are derived from the same primitive  $\boldsymbol{\varepsilon}$  through the same natural mechanism. This guarantees that they are comparable to each other.

---

<sup>7</sup>Since the factor loading matrix  $\Gamma$  is orthogonal, assuming the noise trader's demand for each *original asset*  $\sim$  i.i.d.  $N(0, \sigma^2)$  is equivalent to assuming her demand for each *factor*  $\sim$  i.i.d.  $N(0, \sigma^2)$ .

### 1.2.5 Equilibrium

We say  $\{T, \{\Lambda^{\alpha,i}\}, \{\mathbf{q}^{\alpha,i}\}, \mathbf{p}_T\}$  is an *equilibrium* iff:

- 1) The bundling strategy  $T$  maximizes the originator's payoff  $E_0[\mathbf{1}'\mathbf{p}_T]$ ;
- 2) Given the originator's bundling strategy  $T$ , and the distribution of his exogenous signal  $\mathbf{s}^{\alpha,i}$ , each investor  $(\alpha, i)$ 's information acquisition choice  $\Lambda^{\alpha,i}$  and portfolio choice  $\mathbf{q}^{\alpha,i}$  maximizes his utility (equation 1.1), subject to the capacity constraint (equation 1.2) and the no-forgetting constraint (equation 1.3);

3) Given every investor's portfolio choice  $\{\mathbf{q}^{\alpha,i}\}$ , prices  $\mathbf{p}_T$  clear the market:  $\int_{\alpha,i} \mathbf{q}^{\alpha,i} + \boldsymbol{\varepsilon}_T = \mathbf{1}$  ; and

4) Beliefs are updated using Bayes' law, and expectations are rational; i.e., ex ante beliefs about  $\mathbf{q}^{\alpha,i}$  are consistent with the true distribution of the optimal portfolio.

We say  $\{\{\Lambda^{\alpha,i}\}, \{\mathbf{q}^{\alpha,i}\}, \mathbf{p}_T\}$  is a *subgame equilibrium* induced by a given bundling strategy  $T$  iff conditions 2) to 4) hold.

For tractability, we consider only linear equilibria, in which price(s)  $\mathbf{p}_T$  are linear functions of payoff(s)  $\mathbf{Y}$  and noise trader's effective demand  $\boldsymbol{\varepsilon}_T$ .

### 1.2.6 Summary of Model Setup

The following timeline summarizes the model setup:



Timeline	Originator	Investors	Noise trader
0	Chooses bundling strategy		
1		Decide which information to acquire	
2	Sells assets	Observe signals and choose portfolio	Demands assets
3	consumes payoff	consume payoff	

In principle, the originator can use a continuum of bundling strategies to create two tradable assets. The following proposition shows that they are all equivalent, and thus it suffices to compare two strategies: i)  $T = \mathbf{I}$ , selling the original assets as they are, and ii)  $T = \mathbf{1}'$ , pooling them together into a single asset.

**Proposition 1.2.1** *The investor's information acquisition problem and the originator's payoff are invariant to different bundling strategies that lead to the same tradable asset span.*

The proofs of this and all subsequent propositions are in the appendix.

Intuitively, the noise trader is assumed to demand the linear projection of her "true" demand of original assets onto the tradable asset span. Thus, her effective demand for each risk factor is invariant to different bundling strategies that lead to the same asset span. This further implies that given investors' private signals and demand for tradable assets, the additional information that investors can obtain from prices is also invariant to these different bundling strategies. Each investor's choice set of feasible portfolios of risk factors is also invariant to different bundling strategies that lead to the same asset span. Thus, given his private signals, the investor's optimal portfolio of factors is invariant to these bundling strategies, and

so is his information acquisition problem ex ante. As a result, the total risk premium demanded is also invariant.

For later discussion, for each factor  $j = 1, 2$ , define  $\lambda_{j,T}^a \equiv \int_{\alpha,i} \lambda_{j,T}^{\alpha,i}$ , market average signal precision of factor  $f_j$  induced by bundling strategy  $T$ . And  $\Lambda_T^a \equiv \text{diag}(\lambda_{1,T}^a, \lambda_{2,T}^a)$ . Subscript  $T$  is suppressed if no confusion is caused.

### 1.3 The Upside of Bundling: The Discipline Channel

In this section, to show the upside of asset bundling: the discipline channel, I use the polar case in which total payoff of the assets for sale depends only on a single factor:  $X_1 + X_2 = \sqrt{2}f_1$ . Later in Section 5.4, I show that qualitatively similar results hold as long as the contribution of factor  $f_2$  is sufficiently low.

Consider a mortgage lender, who has originated and wants to sell mortgages on all apartments in New York. To him,  $f_1$  corresponds to common shocks to the prices of all these apartments, and  $f_2$  to shocks specific to the price of a single building whose contribution to the total value of the mortgages for sale is negligible.

As discussed in Proposition 1.2.1, we need to compare only two bundling strategies:  $T = \mathbf{I}$ , selling the original assets as they are, and  $T = \mathbf{1}'$ , pooling them together.

The following proposition further simplifies the analysis, which shows that, to compare the seller's payoffs in the subgames induced by the two bundling strategies respectively, it suffices to compare the corresponding market average signal precision of factor  $f_1$  induced:

**Proposition 1.3.1** *If  $X_1 + X_2 = \sqrt{2}f_1$ , for both  $T = \mathbf{I}$  and  $T = \mathbf{1}'$ , the originator's payoff is  $E_0[\sum X_i] - 2\rho[\frac{1}{\rho^2\sigma^2}(\lambda_{1,T}^a)^2 + \lambda_{1,T}^a]^{-1}$ .*

The originator's payoff depends only on  $\lambda_{1,T}^a$ , since his net supply of  $f_2$  is zero. And his payoff is a strictly increasing function of  $\lambda_{1,T}^a$ , because investors demand a lower risk premium for holding  $f_1$  in equilibrium if, on average, they face less of such risk.

An immediate result is:

**Corollary 1.3.1** *If  $X_1 + X_2 = \sqrt{2}f_1$  and  $K = 1$ , then  $T = \mathbf{I}$  and  $T = \mathbf{1}'$  generate the same payoff to the originator.*

That is, when investors cannot acquire information, the originator is indifferent between bundling the assets and selling them as they are, because the investors' knowledge of  $f_1$  is exogenous.

We now characterize how investors acquire information following the two bundling strategies respectively.

Proposition 1.3.2 shows that pooling the assets induces the originator's desired information acquisition behavior in the investors:

**Proposition 1.3.2** *If  $X_1 + X_2 = \sqrt{2}f_1$ , in the unique subgame equilibrium induced by  $T = \mathbf{1}'$ , every investor learns only about  $f_1$ , regardless of investor type.*

The reason behind this result is intuitive: when the original assets are pooled together, the unique new asset formed has payoff  $Y = X_1 + X_2 = \sqrt{2}f_1$ ; i.e., the diversifiable risk  $f_2$  is washed out. Thus, each investor's portfolio choice problem

is simply how much of  $f_1$ , the non-diversifiable risk to take. Anticipating that, all investors know in advance that they can benefit only from information about  $f_1$ , and thus they only acquire such information in equilibrium. This holds regardless of their expertise.

Since this is the dominant strategy for every investor, the subgame equilibrium induced is unique.

Note that in this subgame equilibrium,  $\lambda_{1,T}^a$ , the market average signal precision of  $f_1$ , reaches the greatest possible level. As a result, bundling strategy  $T = \mathbf{I}$ , selling the original assets as they are, can do no better than pooling them. Indeed, the following proposition indicates that selling the original assets as they are is strictly inferior to pooling them when investors have a big enough capacity  $K$ :

**Proposition 1.3.3** *If  $X_1 + X_2 = \sqrt{2}f_1$ , in the unique subgame equilibrium induced by  $T = \mathbf{I}$ , each investor learns only about one factor, respectively, and*

- 1) *all type 1 investors learn only about  $f_1$ .*
- 2)  *$\exists K_0 < \infty$  such that a positive proportion of type 2 investors learn about  $f_2$  if  $K > K_0$ .*

Although  $f_2$  is diversifiable in aggregation, the loading of each asset on it is generally not zero, like the shock specific to the single building in the example at the beginning of this section. Indeed, given the full rank of the risk-loadings matrix  $\Gamma$ , when assets are sold as they are, investors could hold in their portfolios any amount of any factor. This allows them to profit from their private information about any factor.

As discussed in Section 1.2.3, an investor's preference for early resolution of uncertainty and risk aversion make him specialize in learning about a single factor. In addition, as discussed in Section 1.2.3.2, any given investor's marginal gain of signal precision of a factor from additional input of capacity increases with the capacity already used on that factor. This further strengthens his incentive to specialize in learning about a single factor. So now the question is: which factor would he choose?

Investors face two concerns when choosing which factor to trade and learn about. First, only  $f_1$  is non-diversifiable and carries a premium in equilibrium, which attracts investors to hold and learn about it. Second, investors want to have information about a factor that is better than the market average, so that they can use their superior information to take advantage of others when trading. Therefore, they want to learn about factors studied by fewer people. This is *strategic substitutability* in information acquisition, which attracts each investor to trade and learn about the factor in which he has expertise.

These two concerns work in the same direction for type 1 investors, so they must dedicate all their capacity to  $f_1$  in equilibrium.

However, these concerns work in the opposite direction for type 2 investors, whose expertise is in  $f_2$  instead of  $f_1$ . When assets are not pooled, it might be rational for some of them to learn about  $f_2$ :

Consider a type 2 investor, and assume that everyone other than him learns only about  $f_1$ . Since he has expertise in  $f_2$ , he also has a comparative advantage in learning about it, as discussed in Section 1.2.3. Thus, he is informationally advantageous in  $f_2$  and disadvantageous in  $f_1$ . When everyone has low capacity  $K$ , investors, on average,

still face significant uncertainty about  $f_1$  after learning, and thus the premium carried by  $f_1$  may still be able to attract the type 2 investor to hold it instead of  $f_2$ , and to learn about it to minimize his informational disadvantage. However, when capacity  $K$  becomes large, the premium carried by  $f_1$  decreases (to 0 when  $K \rightarrow \infty$ ), and at the same time, his comparative advantage in learning about  $f_2$  becomes larger and larger. Since others have not yet learned about  $f_2$ , he would prefer to learn about it in order to exploit others with his superior information when trading.

However, from the originator's perspective, the fact that his net supply of  $f_2$  is zero implies that the capacity used to learn about it is a waste of resources. Thus, we have:

**Proposition 1.3.4** *If  $X_1+X_2 = \sqrt{2}f_1$ , in equilibrium the originator chooses  $T = \mathbf{1}'$ , pooling the assets.*

Back to the mortgage lender mentioned at the beginning of this section. He is better off pooling all his mortgages than selling them separately, because pooling prohibits those mortgage buyers who know one particular building better than others from cherry-picking its mortgage, and profiting from information about it. Instead, their attention is drawn to shocks common to all the mortgages for sale, which affects the risk premium. We name this beneficial channel of asset bundling the *discipline channel*.

## 1.4 The Downside of Bundling: The Trade-Restriction Channel and the Specialization-Destruction Channel

What if a bank simultaneously issues and wants to sell off loans of different asset classes, say mortgages and credit cards, that make similar contributions to the total value of the loans? This section shows the bank is better off selling loans of different classes separately.

Formally, we consider the other polar case in which the two factors contribute equally to the total payoff of the original assets:  $X_1 + X_2 = f_1 + f_2$ . Each factor can be thought of as common shocks to a different asset class. Section 1.5.4 shows that qualitatively similar results hold as long as the contributions of the two factors are sufficiently close.

Again, without loss of generality, we consider only two bundling strategies,  $T = \mathbf{I}$ , selling the original assets as they are, and  $T = \mathbf{1}'$ , pooling them together. In this context, pooling the assets creates a new asset with payoff  $Y = X_1 + X_2 = f_1 + f_2$ , which implies that each investor has to hold an equal amount of  $f_1$  and  $f_2$ .

The following proposition states the result of the comparison: pooling the assets is strictly inferior.

**Proposition 1.4.1** *If  $X_1 + X_2 = f_1 + f_2$ , the originator is strictly better off choosing  $T = \mathbf{I}$  instead of  $T = \mathbf{1}'$ .*

Two different economic forces lead to the deficiency of bundling: the *trade-restriction channel* and the *specialization-destruction channel*.

The trade-restriction channel is mechanical and is not related to information acquisition. Thus we illustrate it by shutting down learning; i.e., by considering  $K = 1$ . In Corollary 3.1, when investors cannot acquire information, pooling the assets or not generates the same payoff to the originator. But here, pooling yields a strictly lower payoff. This is because, each investor knows one factor better than the other because of his particular expertise; i.e., from his exogenous signals, and thus wants to trade that factor more aggressively than the other. But this is precluded by the bundling strategy of pooling. The following proposition characterizes investors' expected holding of risk factors, and shows that bundling ( $T = \mathbf{1}'$ ) results in a less efficient allocation of risk factors across investors than not bundling ( $T = \mathbf{I}$ ):

**Proposition 1.4.2** *If  $X_1 + X_2 = f_1 + f_2$  and  $K = 1, \forall i$*

- 1) *If  $T = \mathbf{I}$ , then each type  $i$ 's expected holding of factor  $f_i$  and  $f_{-i}$  are  $\frac{\bar{\lambda} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \lambda}{2})^2}{\frac{\bar{\lambda} + \lambda}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \lambda}{2})^2}$  and  $\frac{\lambda + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \lambda}{2})^2}{\frac{\bar{\lambda} + \lambda}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \lambda}{2})^2}$ , respectively;*
- 2) *If  $T = \mathbf{1}'$ , then each type  $i$ 's expected holding of factor  $f_i$  and  $f_{-i}$  are both 1.*

Intuitively, type  $i$  investors know factor  $f_i$  better than the other type, and are willing to hold  $f_i$  for a lower risk premium. Thus, the originator is better off having them hold more of  $f_i$ . When assets are bundled, an investor is restricted to holding equal amounts of  $f_1$  and  $f_2$ . Since knowledge of factors is symmetric across different types of investor, and since both factors contribute symmetrically to the payoff of the single tradable asset  $Y = X_1 + X_2 = f_1 + f_2$ , each investor in expectation takes an equal share of each factor. However, when assets are not bundled, investors can



freely trade any factor. Since they are risk averse, type  $i$  investors would choose to hold more of  $f_i$ , the factor they know better, and less of  $f_{-i}$ , the factor they know less. This can be seen from  $\frac{\bar{\lambda} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}{\frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2} > 1 > \frac{\underline{\lambda} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}{\frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}$ , as  $\bar{\lambda} > \underline{\lambda}$ .

The trade-restriction channel is driven by the differences in each investor's knowledge of different factors. If each investor knows each factor equally well ( $\bar{\lambda} = \underline{\lambda}$ ), then  $\frac{\bar{\lambda} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}{\frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2} = 1 = \frac{\underline{\lambda} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}{\frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} (\frac{\bar{\lambda} + \underline{\lambda}}{2})^2}$ . Thus bundling or not bundling yields the same allocation of risk factors across investors. The following proposition further confirms this point by showing that if each investor knows each factor equally well, the originator is indifferent between pooling the assets or not:

**Proposition 1.4.3** *If  $X_1 + X_2 = f_1 + f_2$ ,  $K = 1$  and  $\bar{\lambda} = \underline{\lambda}$ , then the originator's payoff is  $E_0[\sum X_i] - 2\rho[\frac{1}{\rho^2 \sigma^2} \underline{\lambda}^2 + \underline{\lambda}]^{-1}$ , whether  $T = \mathbf{I}$  or  $T = \mathbf{1}'$  is chosen.*

The specialization-destruction channel impacts the originator's payoff through its impact on the information acquisition behavior of investors. We now characterize how investors acquire information in the subgames engendered by the two bundling strategies respectively.

The following proposition shows that, if the original assets are sold separately, each investor focuses on his area of expertise:

**Proposition 1.4.4** *If  $X_1 + X_2 = f_1 + f_2$ , in the unique subgame equilibrium induced by  $T = \mathbf{I}$ , each type  $i$  investor learns only about  $f_i$ ,  $\forall i$ .*

Intuitively, when assets are not bundled, investors can freely trade individual factors. As discussed in Proposition 1.3.3, each investor devotes all his capacity to only one factor. Here, both factors play a symmetric role, and carry the same

premium in equilibrium, so the strategic substitutability in information acquisition determines that each investor specializes in his area of expertise, and also determines the uniqueness of subgame equilibrium.

What happens if the assets are pooled,  $T = \mathbf{1}'$ ? The following proposition shows that investors are then induced to spend most of their capacity on the factor in which they have no expertise:

**Proposition 1.4.5** *If  $X_1 + X_2 = f_1 + f_2$ , in the unique subgame equilibrium induced by  $T = \mathbf{1}'$ , each investor tries his best to equalize his knowledge of different factors:*

- 1) *If  $K \geq \bar{\lambda}/\underline{\lambda} > 1$ , then  $\lambda_1^{\alpha,i} = \lambda_2^{\alpha,i} = \sqrt{K\bar{\lambda}\underline{\lambda}}$ ,  $\forall \alpha, i$ ;*
- 2) *If  $1 \leq K < \bar{\lambda}/\underline{\lambda}$ , then  $\lambda_i^{\alpha,i} = \bar{\lambda}$ ,  $\lambda_{-i}^{\alpha,i} = K\underline{\lambda}$ ,  $\forall \alpha, i$ .*

Anticipating that he has to hold equal amounts of each factor, each investor tries his best to equalize his knowledge of different factors through information acquisition, in order to adapt to the trading restriction. Such equalization can be achieved perfectly only if the investor's capacity reaches a threshold  $\bar{\lambda}/\underline{\lambda}$ , which depends on the magnitude of his expertise. When his capacity is below the threshold, he devotes all his capacity to the factor in which he has no expertise. Since this is the dominant strategy for every investor, the subgame equilibrium induced is unique.

When assets are not pooled, each investor focuses on acquiring information in his area of expertise, and his comparative advantage is fully utilized. While if assets are pooled, each investor expends most of his capacity on the factor in which he has no expertise. As a result, similar to the implication of comparative advantage theory in international trade, investors, on average, face more residual uncertainty

after learning about *every* factor when assets are pooled, which leads to a higher risk premium in equilibrium. We call this adverse channel of asset bundling the *specialization-destruction channel*.

As capacity  $K$  increases, investors are more able to adapt their knowledge to the trading restriction, and thus the trade-restriction channel weakens and the specialization-destruction channel strengthens. When  $K < \bar{\lambda}/\underline{\lambda}$ , each investor lacks the capacity to equalize his knowledge of each factor, and both channels are in play. When  $K \geq \bar{\lambda}/\underline{\lambda}$ , investors have enough capacity to achieve perfect equalization of knowledge. In this case, the trade-restriction channel completely disappears, and only the specialization destruction channel plays a role.

Therefore, the banker at the beginning of this section is better off selling the two asset classes separately. This allows the mortgage specialists among the investors to trade mortgages more aggressively relative to credit card loans, and thus induces them to focus on acquiring information about their specialty. This reduces the residual uncertainty faced by investors on average after learning about the mortgages being sold, and thus lowers the risk premium demanded. A symmetric argument applies to credit card specialists.<sup>8</sup>

---

<sup>8</sup>As in Van Nieuwerburgh and Veldkamp (2009), specialization in information acquisition does not imply specialization in factor holding. In equilibrium, type  $i$  investors still want to hold some  $f_{-i}$  for diversification of noise trader risk, which is assumed to be i.i.d. across factors.

## 1.5 Discussion

### 1.5.1 A Generalization: The Optimality of Categorization Strategy

This subsection demonstrates that the economic forces illustrated in the previous two sections carry through to more general environments. We generalize the baseline model to an  $n$ -factor- $n$ -asset setup with  $n$  corresponding types of investors: the payoffs of the original  $n$  assets are  $\mathbf{X} = \Gamma \mathbf{f}$ , where  $\Gamma$  is an  $n$ -by- $n$  orthogonal factor loading matrix. Each type of investors has mass  $1/n$ . Each type  $i$  investor's exogenous signal  $\mathbf{s}^{\alpha,i} \sim N(\mathbf{f}, (\Lambda_0^{(i)})^{-1})$ ,  $\Lambda_0^{(i)} = \text{diag}(\lambda_{0,1}^{(i)}, \dots, \lambda_{0,n}^{(i)})$ ,  $\lambda_{0,i}^{(i)} = \bar{\lambda} > \underline{\lambda} = \lambda_{0,j}^{(i)} \forall j \neq i$ . Each bundling strategy that creates  $1 \leq m \leq n$  tradable assets is uniquely represented by a full rank  $m \times n$  matrix  $T_{m \times n}$  such that  $\mathbf{1}'_m T = \mathbf{1}'_n$ , and that the tradable assets have payoffs  $\mathbf{Y} = T\mathbf{X}$ . Each tradable asset has a supply of 1. Everything else is analogous to the baseline model.

Proposition 1.2.1 shows that the essence of the choice of bundling strategies is the resulting tradable asset span. It is shown in the appendix that this proposition also holds in this generalized setup. In the baseline model, the originator effectively has only two feasible choices of tradable asset spans: either to allow investors to freely trade any of the two factors, or to restrict each investor to trading equal amounts of them. The  $n$ -factor generalized setup significantly expands the originator's set of feasible choices of tradable asset spans to a continuum.

We consider the intermediate case:  $\exists 1 \leq i^* \leq n$ , such that  $\forall i \leq i^*, w_i = w > 0$ ,<sup>9</sup> and  $w_i = 0 \forall i > i^*$ . That is,  $f_1, \dots, f_{i^*}$  are non-diversifiable and play symmetric roles, while  $f_{i^*+1}, \dots, f_n$  are diversifiable. This corresponds to the scenario in which a bank tries to sell loans of  $i^*$  different asset classes, each with many loans and each contributing similarly to the total value of the loans for sale. This nests in the two special cases discussed in the previous two sections, in which  $n = 2$ , and  $i^* = 1$  and 2, respectively.

The aim of this subsection is to establish the optimality of categorization strategy, which corresponds to pooling loans by asset class in the context of securitization and is defined formally as follows:

**Definition 1.5.1** *Categorization strategy is represented by the  $(i^* \times n)$  dimensional matrix  $T$  such that:*

$$T\Gamma = \begin{pmatrix} w\mathbf{I}_{i^*} & \mathbf{0}_{(n-i^*) \times i^*} \end{pmatrix} = \begin{pmatrix} w & 0 & \cdots & 0 & 0 & 0 \\ 0 & w & & \vdots & & \ddots \\ \vdots & & \ddots & 0 & & \\ 0 & \cdots & 0 & w & 0 & 0 \end{pmatrix}$$

.

This strategy creates  $i^*$  tradable assets, such that  $\mathbf{Y} = T\mathbf{X} = T\Gamma\mathbf{f} = (wf_1, wf_2, \dots, wf_{i^*})'$ . It removes all the diversifiable factors asset-by-asset, and each tradable asset takes all the loading of a different non-diversifiable factor.

---

<sup>9</sup>The orthogonality of loading matrix  $\Gamma$  implies  $w = \sqrt{n/i^*}$ .

To establish global payoff optimality, we should ideally compare this strategy with all its opponents. However, the information acquisition problem for each individual investor following an arbitrary bundling strategy  $T$  is intractable. So instead, I prove that categorization strategy achieves a weaker sense of optimality: it implements the capacity allocation and achieves the originator's payoff of an efficiency benchmark. In this hypothetical benchmark, before investors trade assets, the originator could *directly force* them to acquire information in the way that maximizes his payoff, instead of indirectly inducing them to do so by means of asset bundling as we have been discussing. This benchmark is meant to capture the best outcome the originator can achieve by affecting how investors acquire information about his assets.

Formally, this efficiency benchmark is defined as:

$$\max_{\{\Lambda^{\alpha,i}\}} E_0\left[\sum_i p(X_i)\right] = E_0[\mathbf{1}'\mathbf{p}_{\mathbf{I}_n}] \quad (1.4)$$

subject to capacity constraint (1.2) and no-forgetting constraint (1.3)  $\forall \alpha, i$

In other words, this benchmark seeks the solution to the following question: Suppose the originator has to sell the original assets as they are, but could instead *directly assign* a feasible capacity allocation to each investor before he makes his portfolio choice, what is the optimal assignment that maximizes the originator's payoff?

This benchmark is considered for the following reasons.

First, the main focus of this paper is to study how the asset originator induces investors to acquire information in a way that maximizes his profit. This benchmark explicitly highlights such a consideration.

Second, one can also rationalize this approach with a bounded-rationality story: As is the case for economists, it is too complicated for the asset originator in this model to compare the whole continuum of bundling strategies one by one. Thus, he takes a shortcut: He first determines his desired feasible capacity allocation for each investor and then checks whether a simple and commonly used bundling strategy could induce that allocation.

Third, in the intermediate case, this benchmark can be analytically solved, and its unique solution has a clear economic interpretation.

Fourth, the result that categorization strategy implements the benchmark in the intermediate case also has a clear economic interpretation, which combines the intuitions introduced in the previous two sections.

Admittedly, one drawback of this approach is that we cannot analytically rule out the possibility that a particular bundling strategy could achieve a higher payoff for the originator than that of the benchmark.

The following proposition characterizes the efficiency benchmark:

**Proposition 1.5.1** *If  $\exists 1 \leq i^* \leq n$ , such that  $\forall i \leq i^*, w_i = w > 0$ , and  $w_i = 0 \forall i > i^*$ , then the solution to the efficiency benchmark is such that:*

- 1) *each investor of type  $i \leq i^*$  specializes in learning about  $f_i$ ; and*
- 2) *each investor of type  $i > i^*$  specializes in one non-diversifiable factor  $j \leq i^*$ , such that there is an equal mass of investors specializing in learning about each such factor.*

Intuitively, a solution to the efficiency benchmark completely utilizes the expertise of investors on non-diversifiable factors. Since the average precision of private infor-

mation about diversifiable factors does not enter the objective function, no capacity should be spent on them. Having each type  $i > i^*$  investor specializing in exactly one non-diversifiable factor takes advantage of comparative advantage in information acquisition. Last, since all non-diversifiable factors carry equal weight in the objective function, and the objective function is concave in  $\lambda_i^a$ , each non-diversifiable factor should receive the same capacity.

The next proposition gives the conclusion of this subsection: Categorization strategy implements the capacity allocation and the originator's payoff of the efficiency benchmark.

**Proposition 1.5.2** *If  $\exists 1 \leq i^* \leq n$ , such that  $\forall i \leq i^*, w_i = w > 0$ , and  $w_i = 0 \forall i > i^*$ , then in any equilibrium induced by categorization strategy, the aggregate capacity allocation and the resulting originator's payoff are the same as the solution to the efficiency benchmark.*

The intuition of this result combines that developed in the past two sections. The removal of diversifiable factors asset-by-asset prohibits investors from taking them, and deters information acquisition about them. This employs the beneficial discipline channel. The full span on non-diversifiable factors allows investors to take any amount of any of them, and induces perfect specialization in information acquisition about them. This avoids the harmful trade restriction channel and specialization destruction channel.



## 1.5.2 Relation to the Security-Design Literature

Now that the three main economic forces in my model have been illustrated, we are in a good position to discuss its relation to the literature on security design. Bundling loans into different pools and issuing securities backed by them is the defining characteristic of securitization, which plays a significant role in the U.S. economy. Originators and investors typically have asymmetric information (as in my model), raising the concern of adverse selection and moral hazard. In the literature relevant to securitization, the theoretical literature on security design probes how to mitigate such information friction. Section 6 of Gorton and Metrick (2013) provides an excellent survey. In addition to rationalizing the feature of pooling loans by asset class introduced at the beginning of Section 1.1, my model also contributes to this literature in the following aspects as well:

First, for simplification, by assuming a 1-seller-1-buyer setup, existing security design models (e.g. Demarzo and Duffie 1999; Demarzo 2005; Farhi and Tirole 2014) typically abstract from the interaction of heterogeneous security buyers, which is the core of my model. Indeed, the beneficial discipline channel of asset bundling introduced in Section 1.3 is achieved by effectively prohibiting investors with expertise in idiosyncratic factors from using their superior information to exploit others.

Second, a typical security-design model in which the seller has information advantage over the buyer (e.g. Townsend 1979; Dang, Gorton, and Holmstrom 2013; Yang 2013) looks at how to deter or reduce the buyer's costly information acquisition, which is wasteful from a welfare perspective. My model augments this by addressing a complementary question: Given that the buyers ("investors") have decided to ex-

pend a fixed amount of resources ("capacity") to acquire information, how can the seller ("originator") induce them to do so in his preferred way? Indeed, the harmful specialization-destruction channel of asset bundling introduced in Section 1.4 results precisely from the waste of capacity on factors that investors have no expertise in studying.

### 1.5.3 Optimal Bundling Strategies May Not Favor Investors

So far, we have been focusing on maximization of originator's payoff. By definition the respective optimal bundling strategies maximize the originator's utility. They also maximize the market liquidity of the risky assets, in the sense of minimizing the expected total price discount,  $E_0[\sum(Y_i - P_i)]$ , since the fundamental of these assets  $E_0[\sum Y_i] = E_0[\sum X_i]$  is exogenous. A natural follow-up question would be: in general, do the respective optimal bundling strategies that we have identified generate a Pareto improvement relative to the subgame equilibria induced by other bundling strategies? This subsection shows that the answer is no.

Besides the originator, there are two types of agents in the model: the investors and the representative noise trader. Since the noise trader has no well-defined utility function, the following welfare analysis will focus on the investors.

The respective optimal bundling strategies minimize the risks faced by investors and in turn the total risk premium. Since investors are risk averse, one might think that, in general, optimal bundling strategies also maximize investors' utility. However, it turns out that this conjecture is wrong.

Using the case in which total payoff of the assets depends only on a single factor  $f_1$ , the following proposition shows that there could be a conflict of interest between the originator and investors:

**Proposition 1.5.3** *If  $X_1 + X_2 = \sqrt{2}f_1$ , the average expected utility of investors in the subgame induced by  $T = \mathbf{1}'$  is lower than that induced by  $T = \mathbf{I}$ .*

In Section 1.3, it was shown that, if there is only one non-diversifiable factor, the originator prefers pooling the assets ( $T = \mathbf{1}'$ ) to selling the assets as they are ( $T = \mathbf{I}$ ). One might think that pooling the assets prohibits investors from exploiting each other with regard to the diversifiable factor  $f_2$ , and thus, should make them better off. However, it turns out that, on average, investors are actually worse off, for three reasons: First, while pooling the assets reduces the average uncertainty investors face about  $f_1$  after learning, it also reduces the premium they can earn by holding it, and the latter outweighs the former for mean-variance investors<sup>10</sup>. Second, investors profit from the noise trader's demand, because the noise trader always moves price(s) against herself. If investors face less uncertainty about  $f_1$ , their demand for it becomes more elastic, so that the noise trader's demand causes less movement in the price of  $f_1$ , and thus investors profit less from her. Third, pooling restricts the asset span available to the noise trader as well, making her less aggressive in fulfilling her "true" demand for the original assets, and thus, losing less to investors<sup>11</sup>.

---

<sup>10</sup>It is shown in the proof of Proposition 1.9.1 in the appendix that an investor's expected utility is  $\frac{1}{2}E\{(\mathbf{Y} - \mathbf{p})'[Var(\mathbf{Y})]^{-1}(\mathbf{Y} - \mathbf{p})\}$ ; i.e., expected return enters quadratically in the numerator, while variance enters linearly in the denominator.

<sup>11</sup>If  $X_1 + X_2 = f_1 + f_2$ , the third effect works in the opposite direction to the first two, making the impact of the optimal bundling strategy on investors ambiguous.

### 1.5.4 Extrapolation Between the Two Polar Cases

In Sections 1.3 and 1.4, two polar cases were used to show the upside and downside of asset bundling. Recall from Section 1.2 that  $\sum X_i = w_1 f_1 + w_2 f_2$ , where  $w_1 \geq w_2 \geq 0$  are loadings of total payoff of assets on each factor. The polar case in Section 1.3 corresponds to the case of extreme asymmetry in the contributions of the two factors to the total payoff of the assets:  $w_2/w_1 = 0$ , while the polar case in Section 1.4 corresponds to the case of extreme symmetry:  $w_2/w_1 = 1$ . Can we say something about the cases in between:  $0 < w_2/w_1 < 1$ ?

The following proposition shows that, with an increase of the contribution of  $f_2$  relative to that of  $f_1$ ,  $w_2/w_1$ , the originator's payoff in the subgame induced by  $T = \mathbf{1}'$ ; i.e., pooling the assets, changes continuously. So does his payoff in the subgame induced by  $T = \mathbf{I}$ .

**Proposition 1.5.4** *In the baseline model, the originator's payoffs in the subgames induced by  $T = \mathbf{1}'$  and  $T = \mathbf{I}$  both change continuously with  $w_2/w_1$ .*

This proposition establishes the continuity of changes in payoffs with respect to changes in factor loadings. This implies that the results in Section 1.3 and 1.4 are robust to a small perturbation of factor loadings. We have established in Section 1.3 that, if factor  $f_2$  is diversifiable, the originator is strictly better off pooling the assets than selling them as they are when investors have enough capacity to acquire information. The continuity established here implies that the same conclusion holds as long as the contribution of  $f_2$  is sufficiently small. Similarly, we know from Proposition 1.4.1 that the originator is strictly worse off pooling the assets when the two

factors are completely symmetric; i.e.,  $w_1 = w_2$ . And the continuity established here implies that the same conclusion holds as long as the contributions of the two factors are sufficiently close.

## **1.6 Asset Bundling and Corporate Diversification**

Although the theoretical model in this paper is motivated by securitization, the key economic forces highlighted also work in other contexts. By relabeling the asset originator in the model as an entrepreneur who owns several lines of businesses and decides how to set the firm boundaries, the model provides an alternative perspective of corporate diversification. This section discusses this perspective and its relation to existing empirical evidence.

This section is not meant to test the model against existing theories of conglomerate formation, but rather serves two different purposes. First, it shows that although the model is motivated by a feature of securitization, the main economic channels highlighted also apply in other contexts. Second, it shows that by taking the model's perspective a conceptual connection can be built between securitization and conglomerate formation, two issues that are both important in their own right but seemingly remote from each other conceptually.

### **1.6.1 A Market-Side Theory of Corporate Diversification**

"Conglomerate firm production represents more than 50 percent of production in the United States. Given the size of production by conglomerate firms, understanding

the costs and benefits of this form of organization has important implications...For corporate finance, the primary questions about diversification are: 'When does corporate diversification affect firm value?' and 'When diversification adds value, how does it do so?' " (Maskimovic and Phillips 2007)

The literature on corporate diversification took off with the discovery by Lang and Stulz (1994) and Berger and Ofek (1995) of the diversification discount: a typical conglomerate is valued by the stock market at a discount compared with a collection of comparable single-segment firms. This discount represents an economically important puzzle. Consequently, a large number of studies tried to explain the diversification discount, and checks whether the discount is a real empirical phenomenon or an artifact of the measurement process. Maksimovic and Phillips (2013) provide a comprehensive survey of these two strands of literature.

The firm boundary of a conglomerate can be viewed from two complementary perspectives. One is from inside the boundary (the firm side): the different businesses of the firm are managed by the same manager (or team), and the firm's organizational structure affects its market value through its impact on the cash flows generated by its various businesses. The other is from outside the boundary (the market side): the stakeholders of the firm cannot selectively invest in and receive cash flow from any particular business disproportionately relative to the others run by the firm. This view in contrast takes the cash flows as given and looks into how the firm boundary affects how financial market participants acquire relevant information and value the firm.

These two complementary perspectives draw a clear bifurcation of all existing theories of corporate diversification. Most theories take the first view; for example, Maksimovic and Phillips (2002), Stein (1997), Matsusaka and Nanda (2002), Rajan, Servaes and Zingales (2000).<sup>12</sup> Only a very few exceptions take the second view. Krishnaswami and Subramaniam (1999) provide evidence that a spin-off enhances value because it mitigates the information asymmetry in the market about the profitability and operating efficiency of the different divisions of the firm. Hund, Monk and Tice (2010) suggest that, if multiple segment firms have lower uncertainty about mean profitability than single segment firms, rational learning about mean profitability provides an alternative explanation for the diversification discount that does not rely on suboptimal managerial decisions or a poor firm outlook. Vijh (2002) presents one explanation closest to this paper: the increased difficulty facing shareholders investing in shares of conglomerates in their effort to create efficient asset portfolios compared with investing in single-industry firms. However, these theories

---

<sup>12</sup>Maksimovic and Phillips (2002) introduce a neoclassical model that trades off the diseconomies of total firm size due to scarcity of management talent against diminishing returns to scale in each business, and predicts that an entrepreneur with similar productivities in all his businesses would choose the form of conglomerate, and single-segment firms otherwise. In Stein (1997), the headquarter of a financially constrained conglomerate, who has better knowledge of all its businesses than the external financial market, could create value by shifting more funds to its best businesses. And such winner-picking works better if errors in knowledge of different businesses are more correlated. In Matsusaka and Nanda (2002), the cost of internal funding is lower than that of external funding, hence a firm with an internal capital market enjoys the option of avoiding costly external financing when any individual business lacks funds, but is also subject to a more severe overinvesting agency problem. Rajan, Servaes and Zingales (2000) highlight the deficiency of ex post bargaining for profits among segments of a conglomerate when an ex ante division rule cannot be committed to. Ex ante transfer of production factors from a less productive segment to a more productive one on one hand increases allocative efficiency, but on the other hand intensifies the concern of the more productive segment that more profit has to be shared with its deficient counterpart and reduces its incentive in production. For more theories that take the first view, see Maksimovic and Phillips (2013).

generate only a diversification discount, not a premium. This seems to be a salient drawback. To help us understand the pros and cons of corporate diversification from a theoretical perspective, a theory that can generate both diversification discount and premium is a must; As shown in the existing empirical literature, there are circumstances in which a diversification premium is observed; e.g., as documented by Villalonga (2004).

My model suggests an alternative market-side perspective of corporate diversification. It complements the first category by viewing the firm boundary of a conglomerate from a different angle and embellishes the second category by remedying the aforementioned drawback: the discipline channel introduced in Section 1.3 could generate a diversification premium, while the trade restriction-channel and the specialization-destruction channel introduced in Section 1.4 could generate a discount.

## **1.6.2 Empirical Predictions**

My model also yields two empirical predictions that are consistent with existing evidence in the literature. It seems reasonable to assume that when different lines of businesses that a company operates are less similar, it is more likely that investors in the financial market have different expertise in acquiring information about each of them. Thus, my model yields two empirical predictions:

First, for a cross-section of diversified firms *formed for reasons exogenous to my model*, the more similar the businesses that a firm operates, the smaller (greater)



the magnitude of diversification discount (premium) should be observed. This is consistent with:

a) Berger and Ofek's original (1995) paper: "the value loss is smaller when the segments of the diversified firm are in the same two-digit SIC code."

b) In Villalonga (2004): diversified firms in period 1989-96 (defined by the Business Information Tracking Series (BITS), a database that covers the whole U.S. economy at the establishment level), actually trade at a premium on average. But a subsample of firms that are covered and defined by Compustat as conglomerates, which captures purely unrelated diversification, shows a discount.

c) In John and Ofek (1995), concerning firms that increase their focus by selling assets, the average cumulative excess return to the seller on the two days preceding and on the day of the divestiture announcement is positive and is positively related to different measures of increase in focus.

Second, regarding conglomerate formation, a diversified firm is more likely to be formed across similar businesses. This is consistent with:

a) Comment and Jarrell (1995) who show a steady trend toward greater focus during the 1980s, as measured by a revenue-based Herfindahl index, and this is associated with greater shareholder wealth.

b) Two recent empirical papers by Hoberg and Phillips (2012, 2010) that document that multiple-industry firms are more likely to operate in industries that are similar, measured by industry product language overlaps, and that mergers and acquisitions are more likely between firms that use similar product market language.

## 1.7 Conclusion

This paper investigates how a self-interested asset originator can use asset bundling to coordinate the information acquisition of and interaction among investors with different expertise. Three key economic forces novel in the literature are highlighted in the model. The upside of asset bundling is driven by the discipline channel: asset bundling gives investors less incentive to acquire information about risks that are diversified away. As such, the originator successfully induces investors to learn only about risks that cannot be diversified. Since investors face fewer of such risks after learning, they demand a lower risk premium in equilibrium, benefiting the originator. The downside of asset bundling is driven by two different economic forces: Asset bundling mechanically restricts the asset span available to investors, and thus prevents them from holding their respective favorite portfolios. Hence in equilibrium, they demand lower prices to compensate. This is the trade-restriction channel. Asset bundling induces each investor to specialize less in information acquisition about the risk he has expertise in. Since investors' expertise is less utilized, more risks are priced in equilibrium. This is the specialization-destruction channel. These forces rationalize the common practice of bundling loans by asset class in securitization, which is at odds with existing theories based on diversification. The analysis also offers an alternative perspective on conglomerate formation (a form of asset bundling), and the relation to empirical evidence in that context is discussed.

Future research could explore other interesting issues. This paper studies asset bundling in a monopoly setup. It would be interesting to see how the model works in an oligopoly or a monopolistic competition setup. In addition, there are many

other ways that an asset originator can affect how investors with different expertise interact and acquire information; e.g., by designing appropriate auction rules, or choosing what information to reveal to the public, and how. These alternatives would also be interesting to study.

## 1.8 References

1. Admati A. 1985. "A noisy rational expectations equilibrium for multi-asset securities markets", *Econometrica*, 53(3), 629-657.
2. Berger P. and E. Ofek. 1995. "Diversification's effect on firm value", *Journal of Financial Economics*, 37(1), 39-65.
3. Comment R. and G. Jarrell. 1995. "Corporate focus and stock returns", *Journal of Financial Economics*, 37(1), 67-87.
4. Dang T.V., G. Gorton and B. Holmstrom. 2013. "Ignorance, Debt and Financial Crises", working paper
5. Demarzo P. 2005. "The Pooling and Trancing of Securities: A Model of Informed Intermediation", *Review of Financial Studies*, 18(1), 1-35.
6. Demarzo P. and D. Duffie. 1999. "A Liquidity-Based Model of Security Design", *Econometrica* 67, 65-99.
7. Farhi E. and J. Tirole. 2014. "Liquidity Bubbles", *Journal of Economic Theory*, forthcoming.

8. Goldstein I. and L. Yang. 2014. "Information Diversity and Complementarities in Trading and Information Acquisition", *The Journal of Finance*, forthcoming.
9. Gorton G. and A. Metrick. 2013. "Securitization". In: Constantinides G., M. Harris, and R. Stulz. (Ed.), *Handbook of the Economics of Finance*, vol. 2. Elsevier. Chapter 1.
10. Grossman S. and J. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets", *The American Economic Review*, 70(3), 393-408.
11. Hoberg G. and G. Phillips. 2010 "Product market synergies and competition in mergers and acquisitions: a text-based analysis", *Review of Financial Studies*, 23, 3773-3811.
12. ----- . 2012, "Industry choice and product language", working paper.
13. Hund J., D. Monk and S. Tice. 2010. "Uncertainty about average profitability and the diversification discount", *Journal of Financial Economics*, 96, 463-484.
14. John K. and E. Ofek. 1995. "Asset sales and increase in focus", *Journal of Financial Economics* 37, 105-126.
15. Lang L. and R. Stulz. 1994. "Tobin's q, Corporate Diversification, and Firm Performance", *Journal of Political Economy*, 102(6), 1248-1280.
16. Lyandres E. 2007. "Strategic cost of diversification", *Review of Financial Studies*, 20(6): 1901-40.

17. Maksimovic V. and G. Phillips. 2002. "Do Conglomerate Firms Allocate Resources Inefficiently Across Industries? Theory and Evidence", *The Journal of Finance*, 57(2), 721-767.
18. \_\_\_\_\_. 2007 "Conglomerate Firms and Internal Capital Markets". In: Eckbo, B.E. (Ed.), *Handbook of Corporate Finance: Empirical Corporate Finance*, vol. 1. Elsevier/North-Holland, Amsterdam, Chapter 8.
19. \_\_\_\_\_. 2013. "Conglomerate Firms, Internal Capital Markets, and the Theory of the Firm", *Annual Review of Economics*, 5:225-44.
20. Matsusaka J. and V. Nanda. 2002. "Internal Capital Markets and Corporate Refocusing", *Journal of Financial Intermediation*, 11(2), 176-211.
21. Peng L. and W. Xiong. 2006. "Investor Attention, Overconfidence and Category Learning", *Journal of Financial Economics*, 80, 563-602.
22. Rajan R., H. Servaes and L. Zingales. 2000. "The Cost of Diversity: The Diversification Discount and Inefficient Investment", *The Journal of Finance*, 55, 35-80.
23. Shannon C. 1948 "Mathematical theory of communication", *Bell System Technology Journal*, 27, 379-423 and 623-656.
24. Sims C. 2010. "Rational Inattention and Monetary Economics". In: Friedman B. and Woodford M. (Ed.), *Handbook of Monetary Economics*, vol. 3. Elsevier. Chapter 4.

25. Stein J. 1997. "Internal Capital Markets and the Competition for Corporate Resources", *The Journal of Finance*, 52(1), pp. 111-133.
26. Subrahmanyam A. 1991. "A Theory of Trading in Stock Index Futures", *Review of Financial Studies*, 4(1), 17-51.
27. Townsend R. 1979. "Optimal contracts and Competitive Markets with costly State Verification", *Journal of Economic Theory*, 21, 265-293.
28. Vijh A. 2002 The positive announcement-period returns of equity carveouts: asymmetric information or divestiture gains? *Journal of Business*, 75 , 153–190
29. Van Nieuwerburgh S. and L. Veldkamp. 2009. "Information Immobility and the Home Bias Puzzle.", *The Journal of Finance*, 64(3), 1187-1215.
30. ----- . 2010. "Information Acquisition and Under-Diversification", *Review of Economic Studies*, 77(2), 779-805.
31. Villalonga B. 2004. "Diversification Discount or Premium? New Evidence from the Business Information Tracking Series", *The Journal of Finance*, 59, 479-506.
32. Yang M. 2013. "Optimality of Debt under Flexible Information Acquisition", working paper.

## 1.9 Appendix

### 1.9.1 Problems and Propositions in the n-factor setup

The derivations in this subsection are in the n-factor setup introduced in Subsection 1.5.1. The 2-factor setup of the baseline model introduced in Section 1.2 is a special case, and the results here also apply.

Let  $w_i \equiv \Gamma_i'1$ , loading of total asset payoff  $\sum X_i$  on factor  $f_i$ ,  $i = 1, 2, \dots, n$ . Thus,  $\sum X_i = \sum_i w_i f_i$ .

#### Price(s) of tradable assets and the originator's payoff

The posterior of investor  $\alpha$  of type  $i$  about the payoff(s) of tradable assets  $Y = TX$  given his two private signals  $s^{\alpha,i}$  and  $\eta^{\alpha,i}$  is  $N(\boldsymbol{\mu}^{\alpha,i}, (\Omega^{\alpha,i})^{-1})$ , where  $\boldsymbol{\mu}^{\alpha,i} = T(\Lambda^{\alpha,i})^{-1}(\Lambda_0^{(i)}\mathbf{s}^{\alpha,i} + \Lambda_\eta^{\alpha,i}\boldsymbol{\eta}^{\alpha,i})$ , and  $\Omega^{\alpha,i} = (T\Gamma(\Lambda^{\alpha,i})^{-1}\Gamma'T')^{-1}$ .

Define  $\Omega^a \equiv \int_{\alpha,i} \Omega^{\alpha,i}$ , the average precision of all investors' private signal of  $Y$ .

Note that, given the bundling strategy  $T_{m \times n}$  and everyone's information acquisition choice  $\{\Lambda^{\alpha,i}\}$ , the rest of the problem fits in the setup of Admati (1985),<sup>13</sup> which gives the equilibrium prices as a function of asset payoffs  $\mathbf{Y}$  and supply from noise traders  $\boldsymbol{\varepsilon}$ :

---

<sup>13</sup>Investors in Admati (1985) have common priors, while we treat priors as though they were private signals.

$$\begin{aligned}
\mathbf{p}_T &= \mathbf{A}_T + \mathbf{Y} + C_T \boldsymbol{\varepsilon}_T, \text{ where} \\
\mathbf{A}_T &= -\rho \left[ \frac{1}{\rho^2 \sigma^2} \Omega^a (TT') \Omega^a + \Omega^a \right]^{-1} \mathbf{1} \\
C_T &= \rho (\Omega^a)^{-1}
\end{aligned} \tag{1.5}$$

Note that  $(\mathbf{p}_T - \mathbf{A}_T) \sim N(\mathbf{Y}, \Omega_{T,p}^{-1})$ , where  $\Omega_{T,p} = [C_T \text{Var}(\boldsymbol{\varepsilon}_T) C_T']^{-1} = \frac{1}{\rho^2 \sigma^2} \Omega^a (TT') \Omega^a$ . Thus, we have:

**Proposition 1.9.1** *The originator's payoff is  $E_0[\mathbf{1}' \mathbf{p}_T] = \mathbf{E}_0[\mathbf{1}'(\mathbf{A}_T + \mathbf{Y} + C_T \boldsymbol{\varepsilon}_T)] = \mathbf{E}_0[\sum X_i] + \mathbf{1}' \mathbf{A}_T$ , where  $\mathbf{A}_T = -\rho[\Omega_{T,p} + \Omega^a]^{-1} \mathbf{1} = -\rho[\frac{1}{\rho^2 \sigma^2} \Omega^a (TT') \Omega^a + \Omega^a]^{-1} \mathbf{1}$ .*

### Portfolio choice and information acquisition of investors

The following proposition articulates the objective function of an investor:

**Proposition 1.9.2** *The information acquisition problem of investor  $(\alpha, i)$  is*

$$\begin{aligned}
\max_{\Lambda^{\alpha,i}} & Tr[\Omega^{\alpha,i} \Omega_{T,p}^{-1}] + A_T' \Omega^{\alpha,i} A_T \\
s.t. & (1.2) \text{ and } (1.3)
\end{aligned} \tag{1.6}$$

Proof: Observing the price(s)  $\mathbf{p}_T$ , he further updates his belief of  $\mathbf{Y}$ . His posterior becomes  $N(\hat{\boldsymbol{\mu}}^{\alpha,i}, (\hat{\Omega}^{\alpha,i})^{-1})$ , where  $\hat{\boldsymbol{\mu}}^{\alpha,i} = (\hat{\Omega}^{\alpha,i})^{-1}(\Omega^{\alpha,i} \boldsymbol{\mu}^{\alpha,i} + \Omega_{T,p} \mathbf{p}_T)$ , and  $\hat{\Omega}^{\alpha,i} = \Omega^{\alpha,i} + \Omega_{T,p}$ .



From his utility function (1.1), given his capacity choice,  $(\alpha, j)$ 's optimal portfolio choice is

$$\mathbf{q}^{\alpha,i} = \frac{1}{\rho} \hat{\Omega}^{\alpha,i} (\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T) \quad (1.7)$$

Thus, ex ante, his expected utility is

$$U = E_0 \left[ \frac{1}{2} (\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T)' \hat{\Omega}^{\alpha,i} (\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T) \right]$$

He knows the distribution of his exogenous signal  $\mathbf{s}^{\alpha,i}$  ex ante. Conditional on it,  $(\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T)$  is a normal vector, with mean  $-\mathbf{A}_T$ , and variance  $\Omega_{T,p}^{-1} - (\hat{\Omega}^{\alpha,i})^{-1}$ . To derive this variance, note that  $Var(\hat{\boldsymbol{\mu}}^{\alpha,i} | \Lambda_0^i) = T\Gamma(\Lambda_0^i)^{-1}\Gamma'T' - (\hat{\Omega}^{\alpha,i})^{-1}$ ,  $Var(\mathbf{p}_T | \Lambda_0^i) = T\Gamma(\Lambda_0^i)^{-1}\Gamma'T' + \Omega_{T,p}^{-1}$ , and  $cov(\mathbf{p}_T, \hat{\boldsymbol{\mu}}^{\alpha,i} | \Lambda_0^i) = T\Gamma(\Lambda_0^i)^{-1}\Gamma'T'$ .

If a generic random vector  $\mathbf{z} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , then  $E[\mathbf{z}'\mathbf{z}] = \boldsymbol{\mu}'\boldsymbol{\mu} + Tr(\boldsymbol{\Sigma})$ . Hence,

$$\begin{aligned} 2U &= Tr[\hat{\Omega}^{\alpha,i}(\Omega_{T,p}^{-1} - (\hat{\Omega}^{\alpha,i})^{-1})] + A_T' \hat{\Omega}^{\alpha,i} A_T = Tr[\hat{\Omega}^{\alpha,i} \Omega_{T,p}^{-1}] + A_T' \hat{\Omega}^{\alpha,i} A_T - m \\ &= Tr[\Omega^{\alpha,i} \Omega_{T,p}^{-1}] + A_T' \Omega^{\alpha,i} A_T + A_T' \Omega_{T,p} A_T \end{aligned}$$

The last equality is due to  $\hat{\Omega}^{\alpha,i} = \Omega^{\alpha,i} + \Omega_{T,p}$ . Since the third term is exogenous to investor  $(\alpha, i)$ , this proves the proposition. *Q.E.D.*

Now, we can prove Proposition 1.2.1.

**Proof of Proposition 1.2.1:** For 2  $m \times n$  bundling strategy  $T_1$  and  $T_2$ , if they lead to the same asset span, then  $\exists$  a full-rank  $m \times m$  matrix  $M$  such that  $\mathbf{1}'M = \mathbf{1}'$ , and that  $T_2 = MT_1$ , then

$$\mathbf{1}'A_{T_2} = \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} (T_2 T_2') \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} + \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} \right]^{-1} \mathbf{1}$$

$$\begin{aligned}
&= \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} (MT_1 T_1' M') \int_{\alpha,i} (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} \right. \\
&+ \left. \int_{\alpha,i} (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M)^{-1} \right]^{-1} \mathbf{1} \\
&= \mathbf{1}' M \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} (T_1 T_1') \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} + \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \right]^{-1} M' \mathbf{1} \\
&= \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} (T_1 T_1') \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} + \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \right]^{-1} \mathbf{1} = \mathbf{1}' A_T
\end{aligned}$$

So the originator has the same payoff.

Concerning the investor's information acquisition problem,

The first term in (1.6) :

$$\begin{aligned}
Tr[\Omega_{T_2}^{\alpha,i} \Omega_{T_2,p}^{-1}] &= Tr\{(T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} (T_2 T_2') \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} \right. \\
&= Tr\{(MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} \\
&\left. \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} (MT_1 T_1' M') \int_{\alpha,i} (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} \right]^{-1} \right\} \\
&= Tr\{M'^{-1} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} (T_1 T_1') \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \right]^{-1} M'\} \\
&= Tr[M'^{-1} \Omega_{T_1}^{\alpha,i} \Omega_{T_1,p}^{-1} M'] = Tr[\Omega_{T_1}^{\alpha,i} \Omega_{T_1,p}^{-1}].
\end{aligned}$$

And because

$$\begin{aligned}
&\mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} (T_2 T_2') \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} + \int_{\alpha,i} (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} \right]^{-1} \\
&= \mathbf{1}' M \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} (T_1 T_1') \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} + \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \right]^{-1} M' \\
&= \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} (T_1 T_1') \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} + \int_{\alpha,i} (T_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1')^{-1} \right]^{-1} M', \\
&\text{and } \Omega_{T_2}^{\alpha,i} = (T_2 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_2')^{-1} = (MT_1 \Gamma(\Lambda^{\alpha,i})^{-1} \Gamma' T_1' M')^{-1} = M'^{-1} \Omega_{T_1}^{\alpha,i} M^{-1},
\end{aligned}$$

the second term in (1.6)  $A'_{T_2} \Omega_{T_2}^{\alpha,i} A_{T_2} = A'_{T_1} \Omega_{T_1}^{\alpha,i} A_{T_1}$

Therefore, the information acquisition problem is also invariant. This concludes the proof. *Q.E.D.*

**Information Acquisition of Investors and the originator's payoff in the subgame induced by  $T = \mathbf{1}'$**

**Proposition 1.9.3** *In the unique subgame equilibrium induced by  $T = \mathbf{1}'$ , investor  $(\alpha, j)$  minimizes  $\sum_i \{w_i^2(\lambda_i^{\alpha,j})^{-1}\}$ , subject to capacity constraint (1.2) and no-forgetting constraint (1.3).*

Proof: If  $T = \mathbf{1}'$ , then all matrices in (1.6) are scalars, and the investor can only affect  $\Omega^{\alpha,j}$ . The objective function is a decreasing function of  $(\Omega^{\alpha,i})^{-1} = \mathbf{1}'\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma\mathbf{1} = \sum_i \{w_i^2(\lambda_i^{\alpha,j})^{-1}\}$ , and thus the investor chooses to minimize it. This is his dominant strategy, so the subgame equilibrium is unique. *Q.E.D.*

This is intuitive: when all assets are bundled together, each investor faces a one-dimensional problem: to hold how big a proportion of the whole pool,  $\sum_i w_i f_i$ , and thus minimizes his posterior variance of it,  $\sum_i \{w_i^2(\lambda_i^{\alpha,j})^{-1}\}$ , no matter what others are doing.

**Proposition 1.9.4** *In the unique subgame equilibrium induced by  $T = \mathbf{1}'$ , the originator's payoff is*

$$E_0[\sum X_i] - \rho \left\{ \frac{n}{\rho^2 \sigma^2} \left[ \int_{\alpha,j} \left( \sum_i w^2(\lambda_i^{\alpha,j})^{-1} \right)^{-1} \right]^2 + \int_{\alpha,j} \left( \sum_i w^2(\lambda_i^{\alpha,j})^{-1} \right)^{-1} \right\}^{-1}$$

Proof: By Proposition 1.9.1, the originator's payoff is

$$\begin{aligned} \mathbf{E}_0[\sum X_i] + \mathbf{1}'\mathbf{A}_T &= \mathbf{E}_0[\sum X_i] - \rho \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \Omega^a (TT') \Omega^a + \Omega^a \right]^{-1} \mathbf{1} \\ &= \mathbf{E}_0[\sum X_i] - \rho \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,j} (\mathbf{1}'\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma\mathbf{1})^{-1} (\mathbf{1}'\mathbf{1}) \int_{\alpha,j} (\mathbf{1}'\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma\mathbf{1})^{-1} + \int_{\alpha,j} (\mathbf{1}'\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma\mathbf{1})^{-1} \right]^{-1} \\ &= \mathbf{E}_0[\sum X_i] - \rho \left\{ \frac{n}{\rho^2 \sigma^2} \left[ \int_{\alpha,j} \left( \sum_i w^2(\lambda_i^{\alpha,j})^{-1} \right)^{-1} \right]^2 + \int_{\alpha,j} \left( \sum_i w^2(\lambda_i^{\alpha,j})^{-1} \right)^{-1} \right\}^{-1}. \end{aligned} \quad \text{Q.E.D.}$$

**Information Acquisition of Investors and the originator's payoff in the subgame induced by  $T = \mathbf{I}$**

**Proposition 1.9.5** ( $N\mathcal{E}V$ ) *In any equilibrium induced by  $T = \mathbf{I}$ , investor  $(\alpha, j)$  specializes in factor  $i_0$ , where  $i_0 \in \arg \max_i \{L_i \lambda_{0,i}^j\}$ , where*

$$L_i = \left\{ \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2 \right\}^{-1} + \left\{ \rho w_i [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1} \right\}^2$$

Proof: If  $T = \mathbf{I}$ , using the fact that  $\Gamma$  is an orthogonal matrix, the first term in (1.6) is

$$\begin{aligned} & Tr \left\{ (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,j} (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} \int_{\alpha,j} (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} \right]^{-1} \right\} \\ &= Tr \left\{ \Gamma \Lambda^{\alpha,j} \Gamma' \Gamma \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,j} \Lambda^{\alpha,j} \Gamma' \Gamma \int_{\alpha,j} (\Lambda^{\alpha,j}) \right]^{-1} \Gamma' \right\} \\ &= Tr \left\{ \Lambda^{\alpha,j} \left[ \frac{1}{\rho^2 \sigma^2} (\Lambda^a)^2 \right]^{-1} \right\} = \sum_{i=1}^n \left\{ \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2 \right\}^{-1} \lambda_i^{\alpha,j}. \end{aligned}$$

$$\begin{aligned} & \text{Since } A'_T = -\rho \mathbf{1}' \left[ \frac{1}{\rho^2 \sigma^2} \int_{\alpha,j} (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} \int_{\alpha,j} (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} + \int_{\alpha,j} (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} \right]^{-1} \\ &= -\rho \mathbf{1}' \Gamma \left[ \frac{1}{\rho^2 \sigma^2} (\int_{\alpha,j} \Lambda^{\alpha,j})^2 + \int_{\alpha,j} \Lambda^{\alpha,j} \right]^{-1} \Gamma', \text{ and} \\ & \Omega^{\alpha,j} = (\Gamma(\Lambda^{\alpha,j})^{-1} \Gamma')^{-1} = \Gamma'^{-1} \Lambda^{\alpha,j} \Gamma^{-1} = \Gamma \Lambda^{\alpha,j} \Gamma', \end{aligned}$$

the second term in (1.6) is

$$\begin{aligned} & A'_T \Omega^{\alpha,j} A_T \\ &= \rho^2 \mathbf{1}' \Gamma \left[ \frac{1}{\rho^2 \sigma^2} (\int_{\alpha,j} \Lambda^{\alpha,j})^2 + \int_{\alpha,j} \Lambda^{\alpha,j} \right]^{-1} \Lambda^{\alpha,j} \left[ \frac{1}{\rho^2 \sigma^2} (\int_{\alpha,j} \Lambda^{\alpha,j})^2 + \int_{\alpha,j} \Lambda^{\alpha,j} \right]^{-1} \Gamma' \mathbf{1} \\ &= \sum_{i=1}^n \left\{ \rho w_i [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1} \right\}^2 \lambda_i^{\alpha,j}, \text{ since } \Lambda^{\alpha,j} \text{ is diagonal.} \end{aligned}$$

Thus, let  $L_i = \left\{ \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2 \right\}^{-1} + \left\{ \rho w_i [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1} \right\}^2$ . The objective function (1.6) is  $\sum_{i=1}^n L_i \lambda_i^{\alpha,j} = \sum_{i=1}^n L_i \lambda_{0,i}^j y_i$ , where  $y_i \equiv \lambda_i^{\alpha,j} / \lambda_{0,i}^j$ , capacity constraint (1.2) becomes  $\prod_{i=1}^n y_i \leq K$ , and no-forgetting constraint (1.3) becomes  $y_i \geq 1 \forall i$ .

This problem maximizes a sum subject to a product constraint. The second order condition for this problem is positive, meaning the optimum is a corner solution. A simple variational argument can show that investor  $(\alpha, j)$  would specialize in factor  $i_0$ , where  $i_0 \in \arg \max_i \{L_i \lambda_{0,i}^j\}$ . *Q.E.D.*

**Proposition 1.9.6** *In a subgame equilibrium induced by  $T = \mathbf{I}$ , the originator's payoff is*

$$E_0[\sum X_i] - \rho \sum_i w_i^2 [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1}$$

Proof: By Proposition 1.9.1, the originator's payoff is

$$\begin{aligned} \mathbf{E}_0[\sum X_i] + \mathbf{1}' \mathbf{A}_T &= \mathbf{E}_0[\sum X_i] - \rho \mathbf{1}' [\frac{1}{\rho^2 \sigma^2} \Omega^a (T T') \Omega^a + \Omega^a]^{-1} \mathbf{1} \\ &= \mathbf{E}_0[\sum X_i] - \rho \mathbf{1}' [\frac{1}{\rho^2 \sigma^2} [\int_{\alpha,i} (\Gamma(\Lambda^{\alpha,i})^{-1} \Gamma')^{-1}]^2 + \int_{\alpha,i} (\Gamma(\Lambda^{\alpha,i})^{-1} \Gamma')^{-1}]^{-1} \mathbf{1} \\ &= \mathbf{E}_0[\sum X_i] - \rho \mathbf{1}' \Gamma [\frac{1}{\rho^2 \sigma^2} (\int_{\alpha,i} \Lambda^{\alpha,i})^2 + \int_{\alpha,i} \Lambda^{\alpha,i}]^{-1} \Gamma' \mathbf{1} \\ &= \mathbf{E}_0[\sum X_i] - \rho \sum_i w_i^2 [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1}. \quad \text{Q.E.D.} \end{aligned}$$

### Proof of Propositions in Section 1.5.1

By Proposition 1.9.5, the efficiency benchmark (1.4) is

$$\max_{\{\Lambda^{\alpha,j}\}} E_0[\sum X_i] - \rho \sum_i w_i^2 [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1}$$

subject to capacity constraint (1.2) and no-forgetting constraint (1.3)  $\forall \alpha, j$

In the intermediate case we considered:  $\exists 1 \leq i^* \leq n$ , such that  $\forall i \leq i^*, w_i = w > 0$ , and  $w_i = 0 \forall i > i^*$ , this is equivalent to:

$$\min_{\{\Lambda^{\alpha,j}\}} \sum_{i=1}^{i^*} [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1}$$

subject to capacity constraint (1.2) and no-forgetting constraint (1.3)  $\forall \alpha, j$

**Proof of Proposition 1.5.1:**

*Step 1:* In the solution to the efficiency benchmark, no investor should learn any idiosyncratic factor  $i > i^*$ , because the objective function is independent of  $\lambda_i^a$ ,  $\forall i > i^*$ .

*Step 2:* "In the solution to Problem 5.1, each investor  $(\alpha, j)$  would specialize in one factor  $i \leq i^*$ ."

If there is a positive mass  $d$  of type  $j$  investors who learn at least two factors, including factor  $i_1$  and  $i_2$ . Denote the set of these investors  $D$ . And  $\int_D \lambda_{i_1}^{\alpha,j} = K_1 \lambda_{0,i_1}^j$ ,  $\int_D \lambda_{i_2}^{\alpha,j} = K_2 \lambda_{0,i_2}^j$ , where  $K_1 > 1$ ,  $K_2 > 1$ , and  $K_1 K_2 \leq K$ .

If instead we let  $d \in (\frac{K_1-1}{K_1 K_2-1}, \frac{K_1-1}{K_1 K_2-1} (1 - \frac{1}{K_1}))$  proportion of  $(\alpha, j) \in D$  putting all their capacity previously used on factor  $i_1$  and  $i_2$  on  $i_1$  only, and the rest of  $(\alpha, j) \in D$

on  $i_2$  only, the resulting new average precision of private signals of factors are:

$$\begin{aligned}\int_D \tilde{\lambda}_{i_1}^{\alpha,j} &= [dK_1K_2 + 1 - d]\lambda_{0,i_1}^j > K_1\lambda_{0,i_1}^j = \int_D \lambda_{i_1}^{\alpha,j} \\ \int_D \tilde{\lambda}_{i_2}^{\alpha,j} &= [d + (1 - d)K_1K_2]\lambda_{0,i_2}^j > K_2\lambda_{0,i_2}^j = \int_D \lambda_{i_2}^{\alpha,j} \\ \int_D \tilde{\lambda}_i^{\alpha,j} &= \int_D \lambda_i^{\alpha,j} \quad \forall i \neq i_1, i_2, \text{ and} \\ \int_{D^c} \tilde{\lambda}_i^{\alpha,j} &= \int_{D^c} \lambda_i^{\alpha,j} \quad \forall i\end{aligned}$$

This strictly improves the originator's payoff, which is a contradiction.

The first two steps reduce the dimension of the problem from infinity to  $n(i^* - 1)$ . In other words, it suffices to focus on  $\{b_i^j : b_i^j \text{ is the proportion of type } j \text{ investors who specialize in factor } i\}$ .

*Step 3:* "If  $\exists i_0 \neq j_0, i_0, j_0 \leq i^*$ , such that  $b_{i_0}^{j_0} > 0$ , then  $b_{j_0}^k = 0 \quad \forall k \neq j_0$ ."

If  $b_{i_0}^{j_0} = d_1 > 0$  and  $\exists k_0$  s.t.  $b_{j_0}^{k_0} = d_2 > 0$ . Let  $d = \min\{d_1, d_2\}$ . Consider a different allocation  $\{\tilde{b}_i^j : \tilde{b}_{j_0}^{j_0} = b_{j_0}^{j_0} + d, \tilde{b}_{j_0}^{j_0} = d_1 - d, \tilde{b}_{i_0}^{k_0} = b_{i_0}^{k_0} + d, \tilde{b}_{j_0}^{k_0} = d_2 - d, \text{ and } \tilde{b}_i^j = b_i^j \text{ otherwise}\}$ . We have  $\tilde{\lambda}_{j_0}^a = \lambda_{j_0}^a + \frac{dK(\bar{\lambda} - \lambda)}{n} > \lambda_{j_0}^a, \tilde{\lambda}_{i_0}^a = \lambda_{i_0}^a + \frac{dK[(\lambda_{0,i_0}^{k_0})^{-1} - \lambda]}{n} \geq \lambda_{i_0}^a$ , and  $\tilde{\lambda}_i^a = \lambda_i^a \quad \forall i \neq i_0, j_0$ . This yields a strict improvement of the originator's payoff, which is a contradiction.

Let  $f(x) = (x + \frac{1}{\rho^2\sigma^2}x^2)^{-1}, x > 0$ . Then  $f'(x) = -\frac{(\frac{2}{\rho^2\sigma^2}x+1)}{(x+\frac{1}{\rho^2\sigma^2}x^2)^2} < 0$ , and  $f''(x) = \frac{2[\frac{3}{(\rho^2\sigma^2)^2}x^2 + \frac{3}{\rho^2\sigma^2}x+1]}{(x+\frac{1}{\rho^2\sigma^2}x^2)^2} > 0$ .

*Step 4:* "All type  $j \leq i^*$  investors specialize in factor  $j$ ."

Assume otherwise. Then  $\exists i_0 \leq i^*, \exists j_0 \leq i^*$  s.t.  $i_0 \neq j_0$  and  $b_{i_0}^{j_0} > 0$ . This means  $b_{j_0}^{j_0} \leq 1 - b_{i_0}^{j_0} < 1$ . By step 3,  $b_{j_0}^k = 0 \quad \forall k \neq j_0$ , and  $b_{i_0}^{i_0} = 1$ , and thus  $\tilde{\lambda}_{j_0}^a =$

$$\frac{\bar{\lambda} + (n-1)\underline{\lambda}}{n} + \frac{b_{j_0}^{j_0}(K-1)\bar{\lambda}}{n} < \frac{\bar{\lambda} + (n-1)\underline{\lambda}}{n} + \frac{b_{i_0}^{i_0}(K-1)\bar{\lambda}}{n} < \frac{\bar{\lambda} + (n-1)\underline{\lambda}}{n} + \frac{b_{i_0}^{i_0}(K-1)\bar{\lambda} + (K-1)\underline{\lambda} \sum_{k \neq i_0} b_{i_0}^k}{n} = \lambda_{i_0}^a$$

Consider a different allocation  $\{\tilde{b}_i^j : \tilde{b}_{j_0}^{j_0} = b_{j_0}^{j_0} + b_{i_0}^{j_0}, \tilde{b}_{i_0}^{j_0} = 0, \text{ and } \tilde{b}_i^j = b_i^j \text{ otherwise}\}$ .

$$\begin{aligned} & \sum_{i=1}^{i^*} [\tilde{\lambda}_i^a + \frac{1}{\rho^2 \sigma^2} (\tilde{\lambda}_i^a)^2]^{-1} - \sum_{i=1}^{i^*} [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1} \\ &= f[\lambda_{j_0}^a + \frac{b_{i_0}^{j_0}(K-1)\bar{\lambda}}{n}] + f[\lambda_{i_0}^a - \frac{b_{i_0}^{j_0}(K-1)\underline{\lambda}}{n}] - [f(\lambda_{j_0}^a) + f(\lambda_{i_0}^a)] \\ &< f[\lambda_{j_0}^a + \frac{b_{i_0}^{j_0}(K-1)\underline{\lambda}}{n}] + f[\lambda_{i_0}^a - \frac{b_{i_0}^{j_0}(K-1)\underline{\lambda}}{n}] - [f(\lambda_{j_0}^a) + f(\lambda_{i_0}^a)] < 0. \end{aligned}$$

The second to last inequality is due to  $f' < 0$ , and the last inequality is due to  $f'' > 0$ . This means that the proposed alternative allocation strictly improves the originator's payoff, a contradiction.

*Step 5:* " $\forall(\alpha, j)$  s.t.  $j > i^*$ , he specializes in one systematic factor  $i \leq i^*$ , such that there is an equal mass of investors specializing in each systematic factor."

Suppose  $\exists i_1, i_2 \leq i^*$ , s.t.  $\sum_{i=1}^n b_{i_1}^j - \sum_{i=1}^n b_{i_2}^j \equiv 2b > 0$ . By step 4,  $\sum_{i=1}^{i^*} b_{i_1}^j = b_{i_1}^{i_1} = 1 = b_{i_2}^{i_2} = \sum_{i=1}^{i^*} b_{i_2}^j$ , and thus  $\sum_{i=1}^n b_{i_1}^j - \sum_{i=1}^n b_{i_2}^j = \sum_{i=i^*}^n b_{i_1}^j - \sum_{i=i^*}^n b_{i_2}^j = 2b > 0$ . This implies  $\lambda_{i_1}^a - \lambda_{i_2}^a = \frac{2b(K-1)\underline{\lambda}}{n} > 0$ .

Consider an alternative allocation  $\{\tilde{b}_i^j\}$ , such that  $\sum_{i=i^*}^n \tilde{b}_{i_1}^j = \sum_{i=i^*}^n b_{i_1}^j - b$ ,  $\sum_{i=i^*}^n \tilde{b}_{i_2}^j = \sum_{i=i^*}^n b_{i_2}^j + b$ , and  $\tilde{b}_i^j = b_i^j$  otherwise. Then

$$\begin{aligned} & \sum_{i=1}^{i^*} [\tilde{\lambda}_i^a + \frac{1}{\rho^2 \sigma^2} (\tilde{\lambda}_i^a)^2]^{-1} - \sum_{i=1}^{i^*} [\lambda_i^a + \frac{1}{\rho^2 \sigma^2} (\lambda_i^a)^2]^{-1} \\ &= f[\lambda_{i_1}^a - \frac{b(K-1)\underline{\lambda}}{n}] + f[\lambda_{i_2}^a + \frac{b(K-1)\underline{\lambda}}{n}] - [f(\lambda_{i_1}^a) + f(\lambda_{i_2}^a)] < 0. \end{aligned}$$

The inequality is again due to  $f'' > 0$ . This means the proposed alternative allocation strictly improves the originator's payoff, a contradiction. This concludes the proof of the whole proposition. *Q.E.D.*



**Proof of Proposition 1.5.2:** We first prove that the categorization strategy induces the capacity allocation of the solution to the efficiency benchmark.

$$\begin{aligned}
& \text{Note that } (\Omega^{\alpha,j})^{-1} = T\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma'T' \\
& = \begin{pmatrix} w\mathbf{I}_{i^*} & \mathbf{0}_{(n-i^*) \times i^*} \end{pmatrix} \begin{pmatrix} (\lambda_1^{\alpha,j})^{-1} & 0 & \cdots & 0 \\ 0 & (\lambda_2^{\alpha,j})^{-1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & (\lambda_n^{\alpha,j})^{-1} \end{pmatrix} \begin{pmatrix} w\mathbf{I}_{i^*} \\ \mathbf{0}_{(n-i^*) \times i^*} \end{pmatrix} \\
& = w^2 \begin{pmatrix} (\lambda_1^{\alpha,j})^{-1} & 0 & \cdots & 0 \\ 0 & (\lambda_2^{\alpha,j})^{-1} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & (\lambda_{i^*}^{\alpha,j})^{-1} \end{pmatrix} \\
& TT' = T\Gamma\Gamma'T' = \begin{pmatrix} w\mathbf{I}_{i^*} & \mathbf{0}_{(n-i^*) \times i^*} \end{pmatrix} \begin{pmatrix} w\mathbf{I}_{i^*} \\ \mathbf{0}_{(n-i^*) \times i^*} \end{pmatrix} = w^2\mathbf{I}_{i^*}, \\
& \Omega_{T,p} = \frac{1}{\rho^2\sigma^2}\Omega^a TT'\Omega^a = \frac{1}{\rho^2\sigma^2} \int_{\alpha,j} (T\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma'T')^{-1} TT' \int_{\alpha,j} (T\Gamma(\Lambda^{\alpha,j})^{-1}\Gamma'T')^{-1} \\
& = \frac{1}{\rho^2\sigma^2} w^{-2} \begin{pmatrix} \lambda_1^a & 0 & \cdots & 0 \\ 0 & \lambda_2^a & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_{i^*}^a \end{pmatrix} w^2 w \begin{pmatrix} \lambda_1^a & 0 & \cdots & 0 \\ 0 & \lambda_2^a & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_{i^*}^a \end{pmatrix} \\
& = \frac{1}{\rho^2\sigma^2} w^{-2} \begin{pmatrix} (\lambda_1^a)^2 & 0 & \cdots & 0 \\ 0 & (\lambda_2^a)^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & (\lambda_{i^*}^a)^2 \end{pmatrix}, \\
& \text{And } A_T = -\rho[\Omega^a + \Omega_{T,p}]^{-1}\mathbf{1}
\end{aligned}$$

$$\begin{aligned}
&= -\rho[w^{-2} \begin{pmatrix} \lambda_1^a & 0 & \cdots & 0 \\ 0 & \lambda_2^a & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_{i^*}^a \end{pmatrix} + \frac{1}{\rho^2\sigma^2}w^{-2} \begin{pmatrix} (\lambda_1^a)^2 & 0 & \cdots & 0 \\ 0 & (\lambda_2^a)^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & (\lambda_{i^*}^a)^2 \end{pmatrix}]^{-1}\mathbf{1} \\
&= -\rho w^2 \begin{pmatrix} [\lambda_1^a + \frac{1}{\rho^2\sigma^2}(\lambda_1^a)^2]^{-1} & & & \mathbf{0} \\ & \ddots & & \\ & & \ddots & \\ \mathbf{0} & & & [\lambda_{i^*}^a + \frac{1}{\rho^2\sigma^2}(\lambda_{i^*}^a)^2]^{-1} \end{pmatrix} \mathbf{1}
\end{aligned}$$

So the first term in the objective function (1.6) is  $Tr[\Omega^{\alpha,j}\Omega_{T,p}^{-1}]$

$$\begin{aligned}
&= Tr\{w^{-2} \begin{pmatrix} \lambda_1^{\alpha,j} & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \lambda_{i^*}^{\alpha,j} \end{pmatrix} \times [\frac{1}{\rho^2\sigma^2}w^{-2} \begin{pmatrix} (\lambda_1^a)^2 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & (\lambda_{i^*}^a)^2 \end{pmatrix}]^{-1}\} \\
&= \sum_{i=1}^{i^*} \{\frac{1}{\rho^2\sigma^2}(\lambda_i^a)^2\}^{-1} \lambda_i^{\alpha,j}
\end{aligned}$$

And the second term in (1.6) is  $A'_T\Omega^{\alpha,j}A_T$

$$\begin{aligned}
&= \mathbf{1}'(-\rho w^2) \begin{pmatrix} [\lambda_1^a + \frac{1}{\rho^2\sigma^2}(\lambda_1^a)^2]^{-1} & & & \mathbf{0} \\ & \ddots & & \\ & & \ddots & \\ \mathbf{0} & & & [\lambda_{i^*}^a + \frac{1}{\rho^2\sigma^2}(\lambda_{i^*}^a)^2]^{-1} \end{pmatrix} \\
&\times w^{-2} \begin{pmatrix} \lambda_1^{\alpha,j} & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \lambda_{i^*}^{\alpha,j} \end{pmatrix} \\
&\times (-\rho w^2) \begin{pmatrix} [\lambda_1^a + \frac{1}{\rho^2\sigma^2}(\lambda_1^a)^2]^{-1} & & & \mathbf{0} \\ & \ddots & & \\ & & \ddots & \\ \mathbf{0} & & & [\lambda_{i^*}^a + \frac{1}{\rho^2\sigma^2}(\lambda_{i^*}^a)^2]^{-1} \end{pmatrix} \mathbf{1} \\
&= \sum_{i=1}^{i^*} \rho^2 w^2 [\lambda_1^a + \frac{1}{\rho^2\sigma^2}(\lambda_1^a)^2]^{-2} \lambda_i^{\alpha,j}.
\end{aligned}$$

Let  $V_i \equiv \{\frac{1}{\rho^2\sigma^2}(\lambda_i^a)^2\}^{-1} + \rho^2w^2[\lambda_i^a + \frac{1}{\rho^2\sigma^2}(\lambda_i^a)^2]^{-2}$ ,  $i = 1, 2, \dots, i^*$ , and  $V_i \equiv 0$  if  $i > i^*$ . With an argument analogous to the proof of Proposition 1.9.4, it can be proved that investor  $(\alpha, j)$  devotes all his capacity to a single factor  $i_0$ , where  $i_0 \in \arg \max_{i \leq i^*} \{V_i \lambda_{0,i}^j\}$ .

$V_k \equiv 0 \forall k > i^*$  implies that no investor learns about diversifiable factors.  $V_i = V_j > 0 \forall i, j \leq i^*$  implies that  $V_i \lambda_{0,i}^i = V_i \bar{\lambda} > V_i \underline{\lambda} = V_j \lambda_{0,j}^i > 0 = V_k \lambda_{0,k}^i \forall i, j \leq i^*, i \neq j, \forall k > i^*$ , which verifies that every type  $i \leq i^*$  investor would specialize in learning about his own factor  $f_i$ . And since  $V_j \lambda_{0,j}^k = V_i \lambda_{0,i}^k \forall i, j \leq i^*$ , every type  $k > i^*$  investor is indifferent between any two non-diversifiable factors. Thus we have verified that a solution to the efficiency benchmark is a subgame equilibrium.

Now we show that every type  $j \leq i^*$  investor would specialize in his own factor  $j$  in any subgame equilibrium. Suppose among type  $j_1, j_2, \dots, j_m \leq i^*$  investors, there are respectively a strictly positive proportion  $b_{j_1}, b_{j_2}, \dots, b_{j_m}$  who are not learning about their own factors, and WLOG  $b_{j_1} \geq b_{j_2} \geq \dots \geq b_{j_m}$ . By Proposition 1.9.4, for each of such factor  $j_k, k = 1, 2, \dots, m$ , there exists a different factor  $\tilde{j}_k \neq j_k$  such that  $V_{j_k} \bar{\lambda} \leq V_{\tilde{j}_k} \underline{\lambda}$ . This implies  $V_{j_k} \underline{\lambda} < V_{\tilde{j}_k} \underline{\lambda} < V_{\tilde{j}_k} \bar{\lambda}$ , i.e. none of non-type  $j_k$  investors would specialize in factor  $j_k$ . This implies  $\Lambda_{j_1}^a \geq \Lambda_j^a \forall j \leq i^*$ , and thus  $V_{j_1} \geq V_j \forall j \leq i^*$ . Now we have  $V_{j_1} \bar{\lambda} \geq V_{\tilde{j}_k} \bar{\lambda} > V_{\tilde{j}_k} \underline{\lambda}$ , a contradiction.

The mass of type  $j > i^*$  investors learning about each non-diversifiable factor has to be equal. This is because, the factor  $i_0$  that has the strictly least investors specializing in it has  $V_{i_0} > V_i \forall i \neq i_0$ , attracting all type  $j > i^*$  investors to specialize in it in equilibrium, a contradiction.

Lastly, with the expression of  $A_T$  derived above in the proof of this proposition, it is straightforward to verify that the originator's payoff induced by categorization strategy,  $\mathbf{E}_0[\sum X_i] + \mathbf{1}'A_T$ , is identical to that of the efficiency benchmark. This concludes the proof. *Q.E.D.*

## 1.9.2 Propositions in the 2-factor setup

**Proof of Proposition 1.3.1** is straightforward from Proposition 1.9.4 and 9.6, as  $n = 2$ ,  $w_1^2 = 2$  and  $w_2^2 = 0$ .

**Proof of Proposition 1.3.2** is straightforward from Proposition 1.9.3, as  $\sum_i \{w_i^2(\lambda_i^{\alpha,j})^{-1}\} = 2(\lambda_1^{\alpha,j})^{-1}$

**Proof of Proposition 1.3.3:** We use Proposition 1.9.5 to prove this proposition. The result that every investor learns only about one factor directly follows.

If some type 1 investors prefer learning about  $f_2$  to  $f_1$ , then we must have  $L_1\bar{\lambda} \leq L_2\lambda$ . This implies  $L_1\lambda < L_2\bar{\lambda}$ , which means all type 2 investors strictly prefer to learn about  $f_2$ . We have  $\lambda_1^a < \lambda_2^a$ , and thus  $L_1 > L_2$  as  $w_1 > w_2 = 0$ . This implies  $L_1\bar{\lambda} > L_2\lambda$ , a contradiction. So all type 1 investors learn only about  $f_1$ .

Suppose all type 2 investors also learn only about  $f_1$ . Then  $L_2\lambda_{0,2}^2 = \{\frac{1}{\rho^2\sigma^2}(\frac{\bar{\lambda}+\lambda}{2})^2\}^{-1}\bar{\lambda}$ , and  $L_1\lambda_{0,1}^2 \rightarrow 0$  as  $K \rightarrow \infty$ . Thus,  $\exists K_0 < \infty$  such that a positive proportion of type 2 investors learn about  $f_2$  if  $K > K_0$ . *Q.E.D.*

**Proof of Proposition 1.4.4:** Again by Proposition 1.9.5, every investor learns only about one factor.

That every investor specializes in his expertise is a subgame equilibrium, since this implies  $L_1 = L_2$ , and thus  $L_1\bar{\lambda} > L_2\underline{\lambda}$  and  $L_1\underline{\lambda} < L_2\bar{\lambda}$ , justifying each investor's choice.

The equilibrium is unique. Otherwise, say WLOG if some type 2 investors learn about  $f_1$  in equilibrium, then  $L_1\underline{\lambda} \geq L_2\bar{\lambda}$ , which implies  $L_1\bar{\lambda} > L_2\underline{\lambda}$ ; i.e., all type 1 investors learn only about  $f_1$ , and thus  $\lambda_1^a > \lambda_2^a$ , and  $L_1 < L_2$  since  $w_1 = w_2$ . This further implies  $L_1\underline{\lambda} < L_2\bar{\lambda}$ , a contradiction. *Q.E.D.*

**Proof of Proposition 1.4.5:** We first derive the equilibrium capacity allocation.

By Proposition 1.9.3, the optimization problem for investor  $(\alpha, j)$  is  $\min_{\{(\lambda_i^{\alpha,j})^{-1}\}} \sum_i (\lambda_i^{\alpha,j})^{-1}$  s.t.  $(\lambda_i^{\alpha,j})^{-1} \leq (\lambda_{0,i}^j)^{-1}$  and  $\prod_i (\lambda_i^{\alpha,j})^{-1} \geq \frac{1}{K}(\bar{\lambda}\underline{\lambda})^{-1}$ ,  $\forall i$ . The first order condition for this problem is  $1 - \frac{v}{(\lambda_i^{\alpha,j})^{-1}} \prod_l (\lambda_l^{\alpha,j})^{-1} + z_i = 0$ , where  $v$  is the Lagrange multiplier on the capacity constraint and  $z_i$  is the Lagrange multiplier on the no-forgetting constraint for factor  $i$ . We guess and verify that if  $K$  exceeds a cutoff  $K^*$ , the no-forgetting constraint does not bind ( $z_i = 0 \forall i$ ). This implies  $(\lambda_i^{\alpha,j})^{-1} = \frac{v}{K}(\bar{\lambda}\underline{\lambda})^{-1}$ . Taking a product on both sides and imposing the capacity constraint again yields  $v = (\frac{1}{K}(\bar{\lambda}\underline{\lambda})^{-1})^{-\frac{1}{2}}$ , and thus  $(\lambda_i^{\alpha,j})^{-1} = (\frac{1}{K}(\bar{\lambda}\underline{\lambda})^{-1})^{\frac{1}{2}}$  which strictly decreases in  $K$ , verifying the guess.

The cutoff  $K^*$  solves  $\bar{\lambda} = \sqrt{K^*\bar{\lambda}\underline{\lambda}}$ . So  $K^* = \bar{\lambda}/\underline{\lambda}$ . And the result "If  $1 \leq K < \bar{\lambda}/\underline{\lambda}$ , then  $\lambda_i^{\alpha,j} = \begin{cases} \bar{\lambda}, & \text{if } i = j \\ K\underline{\lambda}, & \text{if } i \neq j \end{cases}$ ,  $\forall i, \alpha, j$ " follows from the no-forgetting constraint, which states that if  $z_i > 0$ , then  $\lambda_i^{\alpha,j} = (\lambda_{0,i}^j)^{-1}$ . *Q.E.D.*

**Proof of Proposition 1.4.1:** Let  $g(x) = E_0[\sum X_i] - 2\rho[x + \frac{1}{\rho^2\sigma^2}x^2]^{-1}$ ,  $x > 0$ . It is easy to see that  $g$  is a strictly increasing function of  $x$ .

By Proposition 1.4.2, in the subgame equilibrium induced by  $T = \mathbf{I}$ ,  $\lambda_1^a = \lambda_2^a = \frac{K\bar{\lambda} + \underline{\lambda}}{2}$ . So by Proposition 1.9.6, the originator's payoff is

$$E_0[\sum X_i] - 2\rho[\frac{K\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2\sigma^2}(\frac{K\bar{\lambda} + \underline{\lambda}}{2})^2]^{-1} = g(\frac{K\bar{\lambda} + \underline{\lambda}}{2})$$

By Proposition 1.4.3, in the subgame equilibrium induced by  $T = \mathbf{1}'$ , if  $K \geq \bar{\lambda}/\underline{\lambda}$ ,  $(\lambda_1^{\alpha,j})^{-1} + (\lambda_2^{\alpha,j})^{-1} = 2(\sqrt{K\bar{\lambda}\underline{\lambda}})^{-1} \forall \alpha, j$ . So by Proposition 1.9.4, the originator's payoff is

$$E_0[\sum X_i] - 2\rho[\sqrt{K\bar{\lambda}\underline{\lambda}} + \frac{1}{\rho^2\sigma^2}(\sqrt{K\bar{\lambda}\underline{\lambda}})^2]^{-1} = g(\sqrt{K\bar{\lambda}\underline{\lambda}})$$

If  $K < \bar{\lambda}/\underline{\lambda}$ ,  $(\lambda_1^{\alpha,j})^{-1} + (\lambda_2^{\alpha,j})^{-1} = (K\underline{\lambda})^{-1} + \bar{\lambda}^{-1} \forall \alpha, j$ , so the originator's payoff is

$$E_0[\sum X_i] - 2\rho[(\frac{(K\underline{\lambda})^{-1} + \bar{\lambda}^{-1}}{2})^{-1} + \frac{1}{\rho^2\sigma^2}(\frac{(K\underline{\lambda})^{-1} + \bar{\lambda}^{-1}}{2})^{-2}]^{-1} = g[(\frac{(K\underline{\lambda})^{-1} + \bar{\lambda}^{-1}}{2})^{-1}]$$

By the inequality of arithmetic and geometric means,  $0 < (\frac{(K\underline{\lambda})^{-1} + \bar{\lambda}^{-1}}{2})^{-1} \leq \sqrt{K\bar{\lambda}\underline{\lambda}} < \frac{K\bar{\lambda} + \underline{\lambda}}{2}$ , which concludes the proof. *Q.E.D.*

**Proof of Proposition 1.4.2:** By (1.7), the expected portfolio holdings of investor  $(\alpha, i)$  is  $E[\mathbf{q}_T^{\alpha,i}] = \frac{1}{\rho}\hat{\Omega}^{\alpha,i}E[\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T]$ . As discussed in the proof of Proposition 1.9.2,  $E[\hat{\boldsymbol{\mu}}^{\alpha,i} - \mathbf{p}_T] = -A_T = \rho[\frac{1}{\rho^2\sigma^2}\Omega^a(TT')\Omega^a + \Omega^a]^{-1}\mathbf{1}$ , and  $\hat{\Omega}^{\alpha,i} = \Omega^{\alpha,i} + \frac{1}{\rho^2\sigma^2}\Omega^a(TT')\Omega^a$ .

If  $T = \mathbf{I}$ , by orthogonality of  $\Gamma$ ,  $\Omega^{\alpha,i} = [\Gamma(\Lambda_0^i)^{-1}\Gamma]^{-1} = \Gamma\Lambda_0^i\Gamma'$ .  $\Omega^a = \int_{\alpha,i}\Omega^{\alpha,i} = \Gamma\Lambda_0^a\Gamma'$ , where  $\Lambda_0^a = \int_i\Lambda_0^i = \frac{\bar{\lambda} + \underline{\lambda}}{2}\mathbf{I}$ .  $\frac{1}{\rho^2\sigma^2}\Omega^a(TT')\Omega^a = \frac{1}{\rho^2\sigma^2}(\frac{\bar{\lambda} + \underline{\lambda}}{2})^2\mathbf{I}$ .

Hence,  $E[\mathbf{q}_T^{\alpha,i}] = \Gamma(\Lambda_0^i + \frac{1}{\rho^2\sigma^2}(\frac{\bar{\lambda} + \underline{\lambda}}{2})^2\mathbf{I})\Gamma'\Gamma(\frac{\bar{\lambda} + \underline{\lambda}}{2}\mathbf{I} + \frac{1}{\rho^2\sigma^2}(\frac{\bar{\lambda} + \underline{\lambda}}{2})^2\mathbf{I})^{-1}\Gamma'\mathbf{1}$

$$= \left[ \frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} \left( \frac{\bar{\lambda} + \underline{\lambda}}{2} \right)^2 \right]^{-1} \Gamma (\Lambda_0^i + \frac{1}{\rho^2 \sigma^2} \left( \frac{\bar{\lambda} + \underline{\lambda}}{2} \right)^2 \mathbf{I}) \mathbf{1}.$$

The last equality is due to  $\Gamma' \mathbf{1} = \mathbf{1}$ , because  $w_1 = w_2 = 1$ .

A type  $i$  investor's expected factor holding is  $\Gamma' E[\mathbf{q}_T^{\alpha, i}] = \left[ \frac{\bar{\lambda} + \underline{\lambda}}{2} + \frac{1}{\rho^2 \sigma^2} \left( \frac{\bar{\lambda} + \underline{\lambda}}{2} \right)^2 \right]^{-1} (\Lambda_0^i + \frac{1}{\rho^2 \sigma^2} \left( \frac{\bar{\lambda} + \underline{\lambda}}{2} \right)^2 \mathbf{I}) \mathbf{1}$ . Since  $\Lambda_0^1 = \text{diag}(\bar{\lambda}, \underline{\lambda})$  and  $\Lambda_0^2 = \text{diag}(\underline{\lambda}, \bar{\lambda})$ , this proves the first statement.

If  $T = \mathbf{1}'$ ,  $\forall (\alpha, i)$ ,  $\Omega^{\alpha, i} = (\bar{\lambda} + \underline{\lambda})^{-1} = \Omega^a$ . Thus  $E[\mathbf{q}_T^{\alpha, i}] = 1 \forall (\alpha, i)$ . Since 1 unit of the tradable asset contains 1 unit of each factor, this proves the second statement.

*Q.E.D.*

**Proof of Proposition 1.4.3** is straightforward from the formulas in the proof of Proposition 1.4.1.

**Proof of Proposition 1.5.3:** It is shown in the proof of Proposition 1.9.2 that  $2U^{\alpha, i} = \text{Tr}[\Omega^{\alpha, i} \Omega_{T, p}^{-1}] + A_T' \Omega^{\alpha, i} A_T + A_T' \Omega_{T, p} A_T$ . Thus,

$$\begin{aligned} 2U^a &\equiv \int_{\alpha, i} 2U^{\alpha, i} = \int_{\alpha, i} \text{Tr}[\Omega^{\alpha, i} \Omega_{T, p}^{-1}] + \int_{\alpha, i} (A_T' \Omega^{\alpha, i} A_T + A_T' \Omega_{T, p} A_T) \\ &= \text{Tr}[\Omega^a \Omega_{T, p}^{-1}] + A_T' (\Omega^a + \Omega_{T, p}) A_T \\ &= \text{Tr}[\rho^2 \sigma^2 (TT')^{-1} (\Omega^a)^{-1}] + \rho^2 \mathbf{1}' (\Omega^a + \Omega_{T, p}) \mathbf{1} \end{aligned}$$

Note that the second term =  $-\rho \mathbf{1}' A_T$ . We know from Proposition 1.9.1 that  $\mathbf{1}' A_T$  is the greatest for the originator's favored bundling strategy. Hence, the second term is smaller for  $T = \mathbf{1}'$  than for  $T = \mathbf{I}$ . This comparison corresponds to the first reason in the explanation of Proposition 1.5.3 in the text.

The noise trader's expected loss to the investors is  $E_0[\varepsilon_T'(p_T - Y)] = E_0[\varepsilon_T'(A_T + C_T \varepsilon_T)] = E_0[\varepsilon_T' C_T \varepsilon_T] = \text{Tr}[\rho \sigma^2 (TT')^{-1} (\Omega^a)^{-1}]$ , which is proportional to the first term of  $2U^a$ .

$$\text{If } w_1^2 = 2 \text{ and } w_2^2 = 0, \text{Tr}[\rho^2 \sigma^2 (TT')^{-1} (\Omega^a)^{-1}] = \begin{cases} \rho^2 \sigma^2 [(\lambda_{1,T=I}^a)^{-1} + (\lambda_{1,T=I}^a)^{-1}], & \text{if } T = \mathbf{I} \\ \rho^2 \sigma^2 (\lambda_{1,T=1'}^a)^{-1}, & \text{if } T = \mathbf{1}' \end{cases}$$

By Proposition 1.3.3,  $\lambda_{1,T=1'}^a \geq \lambda_{1,T=I}^a$ . Thus,  $\rho^2 \sigma^2 (\lambda_{1,T=1'}^a)^{-1} \leq \rho^2 \sigma^2 (\lambda_{1,T=I}^a)^{-1} < \rho^2 \sigma^2 [(\lambda_{1,T=I}^a)^{-1} + (\lambda_{1,T=I}^a)^{-1}]$ .

The first inequality corresponds to the second reason in the explanation of Proposition 1.5.3 in the text, and the second inequality to the third reason. *Q.E.D.*

**Proof of Proposition 1.5.4:** First, recall that  $w_1^2 = 2 - w_2^2$ , which monotonically decreases with  $w_2$ . So  $w_2$  increases monotonically with  $w_2/w_1$  in the range we consider:  $0 \leq w_2/w_1 \leq 1$  and  $w_1 > 1$ . Thus, it suffices to focus on change of  $w_2$ .

If  $T = \mathbf{1}'$ , to show that the originator's payoff changes continuously with  $w_2$ , by Proposition 1.9.4, it suffices to show  $[w_1^2 (\lambda_1^{\alpha,i})^{-1} + w_2^2 (\lambda_2^{\alpha,i})^{-1}]$  changes continuously with  $w_2$  for each type.

With an argument analogous to the proof of Proposition 1.4.3, one can show that each investor tries his best to equalize  $w_1^2 (\lambda_1^{\alpha,i})^{-1}$  and  $w_2^2 (\lambda_2^{\alpha,i})^{-1}$  when acquiring information.

For a type 1 investor, if  $w_1^2 \bar{\lambda}^{-1} \geq w_2^2 \underline{\lambda}^{-1}$ , he goes for  $f_1$  first before learning about  $f_2$  :

if  $w_1^2 (K\bar{\lambda})^{-1} > w_2^2 \underline{\lambda}^{-1}$ , then  $\lambda_1^{\alpha,1} = K\bar{\lambda}$ ,  $\lambda_2^{\alpha,1} = \underline{\lambda}$ ;

if  $w_1^2 (K\bar{\lambda})^{-1} \leq w_2^2 \underline{\lambda}^{-1}$ , then  $w_1^2 (\lambda_1^{\alpha,1})^{-1} = w_2^2 (\lambda_2^{\alpha,1})^{-1} = w_1 w_2 \sqrt{(K\bar{\lambda}\underline{\lambda})^{-1}}$ ;

If  $w_1^2 \bar{\lambda}^{-1} < w_2^2 \underline{\lambda}^{-1}$ , he goes for  $f_2$  first before learning about  $f_1$  :

if  $w_1^2 \bar{\lambda}^{-1} < w_2^2 (K\underline{\lambda})^{-1}$ , then  $\lambda_1^{\alpha,1} = \bar{\lambda}$ ,  $\lambda_2^{\alpha,1} = K\underline{\lambda}$ ;

if  $w_1^2 \bar{\lambda}^{-1} \geq w_2^2 (K\underline{\lambda})^{-1}$ , then  $w_1^2 (\lambda_1^{\alpha,1})^{-1} = w_2^2 (\lambda_2^{\alpha,1})^{-1} = w_1 w_2 \sqrt{(K\bar{\lambda}\underline{\lambda})^{-1}}$ .



$$\text{Therefore, } w_1^2(\lambda_1^{\alpha,1})^{-1} + w_2^2(\lambda_2^{\alpha,1})^{-1} = \begin{cases} w_1^2(K\bar{\lambda})^{-1} + w_2^2\bar{\lambda}^{-1}, & \text{if } \frac{w_2^2}{w_1^2} < \frac{\lambda}{K\bar{\lambda}} \\ 2w_1w_2\sqrt{(K\bar{\lambda}\lambda)^{-1}}, & \text{if } \frac{\lambda}{K\bar{\lambda}} \leq \frac{w_2^2}{w_1^2} \leq \frac{K\lambda}{\bar{\lambda}} \\ w_1^2\bar{\lambda}^{-1} + w_2^2(K\lambda)^{-1}, & \text{if } \frac{w_2^2}{w_1^2} > \frac{K\lambda}{\bar{\lambda}} \end{cases},$$

Since  $\frac{w_2^2}{w_1^2}$  strictly increases with  $w_2$  in the range of interest,  $w_1^2(\lambda_1^{\alpha,1})^{-1} + w_2^2(\lambda_2^{\alpha,1})^{-1}$  is continuous in  $w_2$  in each segment, and its value coincides at the two thresholds.

For a type 2 investor, since  $w_1^2\bar{\lambda}^{-1} > w_2^2\bar{\lambda}^{-1}$ , he always goes for  $f_1$  first before learning about  $f_2$ :

if  $w_1^2(K\lambda)^{-1} > w_2^2\bar{\lambda}^{-1}$ , then  $\lambda_2^{\alpha,2} = K\lambda$ ,  $\lambda_2^{\alpha,2} = \bar{\lambda}$ ;

if  $w_1^2(K\lambda)^{-1} > w_2^2\bar{\lambda}^{-1}$ , then  $w_1^2(\lambda_1^{\alpha,2})^{-1} = w_2^2(\lambda_2^{\alpha,2})^{-1} = w_1w_2\sqrt{(K\bar{\lambda}\lambda)^{-1}}$ .

$$\text{Therefore, } w_1^2(\lambda_1^{\alpha,1})^{-1} + w_2^2(\lambda_2^{\alpha,1})^{-1} = \begin{cases} w_1^2(K\bar{\lambda})^{-1} + w_2^2\bar{\lambda}^{-1}, & \text{if } \frac{w_2^2}{w_1^2} < \frac{\bar{\lambda}}{K\lambda} \\ 2w_1w_2\sqrt{(K\bar{\lambda}\lambda)^{-1}}, & \text{if } \frac{w_2^2}{w_1^2} \geq \frac{\bar{\lambda}}{K\lambda} \end{cases}, \text{ which}$$

is again continuous in  $w_2$  in each segment, and its value coincides at the threshold.

Thus, we have proved that if  $T = \mathbf{1}'$ , to show that the originator's payoff changes continuously with  $w_2$ .

Now we prove that the same conclusion holds if  $T = \mathbf{I}$ .

As in Proposition 1.9.5, let  $L_i = \{\frac{1}{\rho^2\sigma^2}(\lambda_i^a)^2\}^{-1} + \{\rho w_i[\lambda_i^a + \frac{1}{\rho^2\sigma^2}(\lambda_i^a)^2]^{-1}\}^2$ ,  $i = 1, 2$ .

Since  $w_1 \geq w_2 \geq 0$ , by an argument analogous to the proof of Proposition 1.4.3, all type 1 investors must learn only about  $f_1$  in equilibrium.

Each type 2 investor either learns about only  $f_1$  or only  $f_2$ . Let  $b \in [0, 1]$  denote the proportion of them who learn  $f_2$ . It suffices to show that  $b$  changes continuously with  $w_2$ .

Then  $\lambda_1^a = \frac{K(\bar{\lambda}+\lambda)}{2} - b\frac{(K-1)\lambda}{2}$ , which strictly decreases in  $b$ , and  $\lambda_2^a = \frac{\bar{\lambda}+\lambda}{2} + b\frac{(K-1)\bar{\lambda}}{2}$ , which strictly increases in  $b$ .

$L_1, L_2$  can be viewed as functions of  $b, w_2$  and  $K$ ,  $L_1(b, w_2, K)$  strictly increases in  $b$ , decreases in  $w_2$  and  $K$ .  $L_2(b, w_2, K)$  strictly decreases in  $b$ , increases in  $w_2$ , and decreases in  $K$ .

$\forall K \geq 1$ , it is easy to verify that  $\underline{\lambda}L_1(1, 0, K) > \bar{\lambda}L_2(1, 0, K)$ , and  $\underline{\lambda}L_1(1, 1, K) < \bar{\lambda}L_2(1, 1, K)$ . Hence  $\exists 0 < \bar{w} < 1$ , such that  $\underline{\lambda}L_1(1, \bar{w}, K) = \bar{\lambda}L_2(1, \bar{w}, K)$ .

If  $w_2 > \bar{w}$ ,  $\underline{\lambda}L_1(1, w_2, K) < \bar{\lambda}L_2(1, w_2, K)$ , by Proposition 1.9.5, each type 2 investor strictly prefers to learn about  $f_2$ , thus  $b = 1$

It is easy to see  $\forall K \geq 1$ ,  $\underline{\lambda}L_1(0, 1, K) < \bar{\lambda}L_2(0, 1, K)$ .

Since  $\underline{\lambda}L_1(0, 0, \infty) < \bar{\lambda}L_2(0, 0, \infty)$  and  $\underline{\lambda}L_1(0, 0, 1) > \bar{\lambda}L_2(0, 0, 1)$ ,  $\exists \bar{K} \in (1, \infty)$  s.t.  $\underline{\lambda}L_1(0, 0, \bar{K}) = \bar{\lambda}L_2(0, 0, \bar{K})$ .

If  $K < \bar{K}$ , we have  $\underline{\lambda}L_1(0, 0, K) > \bar{\lambda}L_2(0, 0, K)$  and  $\underline{\lambda}L_1(0, 1, K) < \bar{\lambda}L_2(0, 1, K)$ . Hence  $\exists 0 < \underline{w} < 1$ , such that  $\underline{\lambda}L_1(0, \underline{w}, K) = \bar{\lambda}L_2(0, \underline{w}, K)$ .

Thus, if  $K < \bar{K}$  and  $w_2 < \underline{w}$ ,  $\underline{\lambda}L_1(0, w_2, K) > \bar{\lambda}L_2(0, w_2, K)$ , by Proposition 1.9.5, each type 2 investor strictly prefers to learn about  $f_1$ , thus  $b = 0$

If  $K < \bar{K}$  and  $\bar{w} \geq w_2 \geq \underline{w}$  or if  $K \geq \bar{K}$  and  $w_2 \leq \bar{w}$ , we have  $\underline{\lambda}L_1(0, w_2, K) \leq \bar{\lambda}L_2(0, w_2, K)$  and  $\underline{\lambda}L_1(1, w_2, K) \geq \bar{\lambda}L_2(1, w_2, K)$ . Hence  $\exists \bar{b}(w_2, K) \in [0, 1]$  s.t.  $\underline{\lambda}L_1(\bar{b}, w_2, K) = \bar{\lambda}L_2(\bar{b}, w_2, K)$ . By Proposition 1.9.5, each type 2 investor is indifferent between learning about  $f_1$  and  $f_2$ , thus a proportion  $\bar{b}(w_2, K)$  of them learning about  $f_2$  constitutes an equilibrium.  $\bar{b}(w_2, K)$  increases continuously in  $w_2$   $\bar{b}(\bar{w}, K) = 1$ , and  $\bar{b}(\underline{w}, K) = 0$  if  $K < \bar{K}$ .

To summarize, if  $K \geq \bar{K}$ , the equilibrium  $b = \begin{cases} \bar{b}(w_2, K), & \text{if } w_2 \leq \bar{w} \\ 1, & \text{if } w_2 > \bar{w} \end{cases}$ . If  $K < \bar{K}$ ,

the equilibrium  $b = \begin{cases} 0, & \text{if } w_2 < \underline{w} \\ \bar{b}(w_2, K), & \text{if } \underline{w} \leq w_2 \leq \bar{w} \\ 1, & \text{if } w_2 > \bar{w} \end{cases}$ .

In both cases,  $b$  varies continuously with  $w_2$ . This concludes the proof. *Q.E.D.*

This document serves as a template to demonstrate how to use the `pthesis` documentclass for a Princeton University Ph.D. Dissertation. Some of the requirements for a master's thesis or an undergraduate thesis are different, especially the text on the title page, so you will need to make some modifications to use this template for those purposes.

This template is setup to easily make a few different versions of your dissertation. The version you will print and have bound should generally be single spaced (and single-sided), and not contain any hyperlinks. The electronic version that you submit to your readers to review and to be published by ProQuest will be double spaced, and will contain PDF features such as bookmarks for each section and internal links to citations, footnotes, and other internal references.

During the writing process, you may want to disable some of the frontmatter (list of tables, list of figures, acknowledgements, and maybe even the table of contents. I have not tested this template with equations or a list of symbols, but those are available.

## Chapter 2

# The Middle Income Trap, Branching Deregulation, and Political Influence

### 2.1 Introduction

The middle income trap is an important economic issue in reality.<sup>1</sup> South American countries that were already pretty wealthy one century ago have experienced substantial economic stagnation after the Second World War. This is a perfect example for the middle income trap. Namely, when the economy achieves a certain level of development, it stops growing further. On the contrary, the Four Asian Tigers (Tai-

---

<sup>1</sup>According to Wikipedia, “the middle income trap is a theorized economic development situation, where a country which attains a certain income (due to given advantages) will get stuck at that level”.

wan, Hong Kong, Singapore and South Korea) which were much poorer than these South American countries one century ago have achieved amazing economic growth after the Second World War, and have roughly the same level of GDP per capita as most developed economies.<sup>2</sup> Then, a natural question to ask is why these two types of economies have achieved different economic outcomes in the past century. More specifically, why do some economies have escaped from the middle income trap, while others have not?

Economists have studied factors that cause the middle income trap from various perspectives. For instance, the technological perspective of the middle income trap argues that underdeveloped economies can use imitation and technological transfers to boost their economic growth at the early stage of the development, since they are far away from the world technological frontier. However, when their economic development achieves a certain level, developing countries have to innovate by themselves. This may be hard for them, since they are used to imitating and receiving technology transfers from developed countries. This eventually causes the middle income trap. Another view is from the demographic perspective. Namely, countries that receive “demographic dividend” (i.e., a higher working age ratio) are more easily to get out of the middle income trap (e.g., Bloom and Williamson, 1998; Bloom and Canning, 2004). Although these views do seem plausible, they all overlook one important factor for economic development, which is the role played by the government.

In this paper, we provide a new explanation for the middle income trap, and this new view is from a political economy perspective. In our framework, the government

---

<sup>2</sup>The GDP per capita of these four economies is all above 30,000 US dollars in 2013 (PPP based).

is subject to the influence of political contributions. As a result, it implements a credit market policy to maximize a weighted average of social welfare and the political contributions. We show that only when the economy achieves a certain level of development (i.e., the productivity of firms in the real sector is high enough) does the government have incentives to design a distorted policy. This distorted policy reduces social welfare and hinders further development of the economy. In other words, political incumbents have a greater incentive to distort the resource allocation and receive rents only when the economic development achieves a certain level. The reason is simple: they can extract few rents when the economy is poor. However, the greater incentive to distort the resource allocation in the market prevents the economy from further development, which creates the middle income trap in our story.

Our explanation has theoretical and empirical support in the literature. First, a vast literature on economic institutions (Buchanan and Tullock, 1963; North, 1981 and 1990; and DeLong and Shleifer, 1993; Jayaratne and Strahan (1996)) has emphasized the advantages of limited government and unregulated credit market for economic development. However, none of them emphasizes the importance of distorted economic institutions that are *endogenously* designed by the government for the middle income trap. Our study bridges this gap by showing the exact *reason* why better economic performance of the economy might hinder its further development by incentivizing political incumbents to implement distorted policies. Second, existing explanations for the middle income trap mentioned above cannot rationalize why countries with similar technological and demographic background have achieved

different outcomes to escape the middle income trap. For instance, Indonesia and Malaysia which had similar technological and demographic background to Singapore and Hong Kong in 1960s have not got out of the middle income trap. Moreover, key differences between these two types of economies are the cleanness of the government and the regulation in the credit market,<sup>3</sup> and they are the key institutions (argued in this paper) for the economy to escape from the middle income trap. In other words, Singapore and Hong Kong have escaped from the middle income trap, because compared with Indonesia and Malaysia, their economic policies and regulation in the financial market are much less distorted even when the economies achieved a certain level of development. Our study shows that intensive monitoring on the government is needed to prevent it from receiving political contributions and designing policies that favor special interest groups. Finally, using cross-country data on economic performance and institutions, Aiyar *et al.* (2013) did find that countries that reduce government involvement in the economy, deregulate their credit markets, and eliminate corruption are more likely to get out of the middle income trap. This empirical finding directly support the argument advocated in this paper.

We use a particular scenario which is the branching deregulation of U.S. banking industry from 1970's to 1990's to both theoretically and empirically elaborate on our argument for the middle income trap. Economic literature has shown that branching regulation of U.S. banking industry had impeded the country's financial development

---

<sup>3</sup>The Corruption Perceptions Index (from Transparency International) show that Singapore and Hong Kong were ranked the 5th. and the 15th. in the world in terms of government cleanness in 2013, while Indonesia and Malaysia were ranked the 53th. and the 114th in 2013. It is also well known that the regulation of financial markets is at the minimum level in Singapore and Hong Kong.

and economic growth. For instance, in their influential paper, Jayaratne and Strahan (1996) showed that the rates of real per capita growth in income and output increase significantly following intrastate branch reform, due to improvements in the quality of bank lending. Ten years later, Cetorelli and Strahan (2006) provided evidence that a concentrated banking sector increases the difficulty for potential entrants in non-financial sectors to obtain credits. However, U.S. had experienced a long history of branching regulation, which is hard to be explained by the public interest theory (Kroszner, 2001). Historical evidence showed that with few exceptions, state banks were not allowed to set up multi branches within their own states and in other states, due to opposition from unit banks (White, 2000). The Macfadden Act enacted in 1927 prohibited interstate branching by allowing each national bank to branch only within the state in which it is situated, until it was repealed by the Riegle-Neal Act in 1994.

Political economy story has been proved to square well with the history of financial development. For example, Rajan and Zingales (2003) proposed an interest group theory to explain the non-monotonic development path of financial markets, and related it to non-financial sectors. Kroszner and Strahan (1999) provided evidence that interest group factors related to the relative strength of potential winners (large banks and small, bank-dependent firms) and losers (small banks and the rival insurance firms) can explain the timing of branching deregulation across states and congressional voting on interstate branching deregulation. Nevertheless, these papers have provided neither theoretical models nor empirical evidence suggesting the *exact channels* through which political interest affects branching deregulation decisions.



In this paper, following Grossman and Helpman (1994), we build a common-agency model to highlight lobbying as a channel through which relevant interest groups affect the deregulation decision. In our model, the incumbent bank can make a political contribution to the government for the policy it favors, since it can overcome the free-ride problem faced by firms in the production sector, who are its borrowers. Due to the branching regulation, the monopoly power of the incumbent bank leads to a high interest rate and a severe credit rationing, deterring credit to many firms with projects that have positive Net Present Value (NPV). In return for the monopoly profit the incumbent bank enjoys, it turns in a higher contribution to the government. When deciding whether or not to allow outside banks to enter and compete with the incumbent bank, the government trades off the higher political contribution from the incumbent bank and a higher social welfare in a competitive banking regime, which is due to a lower interest rate and alleviated credit rationing. When the expected NPV from the production sector is high, which is exactly when the competitive banking regime is more desired in the utilitarian view, the incumbent bank could extract a higher profit from its borrowers, and thus has a greater incentive to keep its monopoly power by contributing more to the government. This makes deregulation less likely when the economy is in boom. Therefore, our model predicts that branching regulation is more likely to be removed during recessions, and when the institutional financial friction is low. The former prediction is consistent with the empirical result presented in Freeman (2002).

We then derive other empirical predictions of the model and provide supporting evidence. In addition to the result discussed above, the model predicts another three

results. First, the bank's profit varies more than the real Gross Domestic Product (GDP) with business cycle. This is because the monopoly bank can extract disproportionately more profit from its borrowers, when the NPV of the borrowers' projects increases. Second, the Net Interest Margin (NIM) decreases and banks receive shrinking profit from loans after deregulation, and more so for small banks. These are direct implications from the comparison between a monopolized banking sector and a competitive one. Finally, as implied by both Grossman-Helpman (1994) and our model, the campaign contribution and lobbying expenditure from the commercial banking sector should be highly positively correlated with its profit. Furthermore, the elasticity of the political contribution with respect to the commercial banking sector's profit should be close to one, as implied by the model. The first and the third predictions are unique empirical predictions of the model, and are directly linked to the channel through which the political economy component of our framework affects the deregulation policy.

Using data collected from the Wharton Research Data Services (WRDS) and the Federal Election Committee (FEC), we show that all the above three empirical predictions gain support from the data. In particular, the data from the WRDS show that the variance of the banking sector's profit at the state level is five times more as volatile as the variance of the state-level GDP. Furthermore, the manually collected data from the FEC reveal that the elasticity of the campaign contribution and lobbying expenditure with respect to the commercial banking sector's profit at the state level is between 0.91 and 0.98, which is extremely close to one. These findings constitute two contributions of this paper relative to the existing literature.

First, our paper provides micro-level (i.e., at the level of Political Action Committees (PACs)) evidence on political contributions and their relationship with the profit of the banking sector. To our best knowledge, this is the first paper that *directly* tests the prediction of Grossman and Helpman (1994) using *micro-level* data.<sup>4</sup> Second, our paper tests and finds support for *each step of the channel* through which political economy affects economic policies. Previous research (e.g., Kroszner and Stratmann (1998) and Kroszner and Strahan (1999)) only provided evidence to support the final predictions of some existing theories or models.

The most important contribution of this paper is the new explanation for the middle income trap phenomenon. We think that this new argument will attract much more attention in the new future. In fact, China's recent experience is a perfect example to show the validity of our argument. After China's opening up to the world market in 1978, the economic reforms were undistorted at the initial stage. Rural Part of China had achieved fast growth from early 1980s to the mid-1990s. Both private enterprises and Township and Village Enterprises (TVEs) had achieved amazing development within the same period of time. However, economic reforms were reversed from the mid-1990s. First, capital and bank loans were generously poured into State-Owned Enterprises (SOEs) which were shown to have lower productivity than private firms (Brandt, Van Biesebroeck, and Zhang (2012) and Khandelwal, Schott, and Wei (2013)). At the meantime, private firms have suffered severely from the lack of credit from commercial banks. The fast development of the underground banking sector in the eastern coastal area of China reflected the

---

<sup>4</sup>Goldberg and Maggi (1999) used industry-level data on political contributions to test several predictions of Grossman and Helpman (1994).

severity of this issue.<sup>5</sup> Second, the central government of China has put forward a slogan of Guojin Mintui (the state advances, the private sector retreats) since 1997. This policy distortion has substantially reduced the profitability of private firms which have higher productivity than the SOEs and benefited the SOEs as well as the government to a large extent. As a result, Brandt, Tombe, and Zhu (2013) found that the loss in aggregate non-agricultural Total Factor Productivity (TFP) due to factor misallocation in China had declined from 1985 to the mid-1990s, and it has increased appreciably from the mid-1990s. They further showed that this reversal can be attributed almost exclusively to increasing misallocation of capital between state and non-state sectors within provinces. In 1980, China's GDP per capita (PPP based) was only about 3.43% percent of the US. However, this ratio increased to 10.14% percent of the US.<sup>6</sup> Therefore, this finding mirrors our argument that the Chinese government did not distort the resource allocation in the economy until the time when the economic development has achieved a certain level.

The remainder of the paper is organized as follows. Section two presents the model. Section three provides evidence to support empirical predictions of the model. Section four concludes.

## 2.2 The Model

In this section, we present a model featuring the tradeoff between the higher social welfare and the higher political contribution faced by the government. We further

---

<sup>5</sup>see Haung (2008) for more details.

<sup>6</sup>The link is <http://www.nationmaster.com/country-info/stats/Economy/GDP-per-capita/PPP>.

derive the main prediction of the model that branching deregulation is more likely to happen when the economy is in recession.

### 2.2.1 Environment

There are three types of players in the state's economy: firms in the product market, the commercial bank and the state government. First, there is a continuum of heterogenous firms in each state, and the mass of them is normalized to one. Firms differ in their profitability level,  $R$ , which has a distribution on  $[\underline{R}, \bar{R}]$ . Second, there is only one commercial bank within each state that can lend money to firms in that state. Finally, the state government is subject to political influence. As a result, it maximizes a weighted average of the total profit of firms in that state and the commercial bank's profit on the fashion à la Grossman and Helpman (1994).

There are three periods in the model. In the first period, the government decides whether or not to liberalize its banking sector. In the second period, the commercial bank in that state chooses an optimal interest rate to maximize its expected profit. In the third period, firms apply for loans, and the commercial bank screens the firms' proposed projects. As a result, the commercial bank learns whether the profitability of each project (i.e.,  $R$ ) is above a certain threshold which is the minimum requirement for firms to receive loans from the bank. In the end, only firms with highly profitable projects obtain loans from the commercial bank and make investment. For simplicity, we assume that the discount factor is one, and the net risk-free interest rate between periods is zero.

We do not model the commercial bank's cost to acquire information about the profitability of the projects explicitly, although this cost implicitly determines some equilibrium outcomes. First, it is probably true in reality that commercial banks do not *fully* know the profitability of each project it funds, since the cost to acquire the *exact* information is too high. This correspond to the assumption of our model that the incumbent bank cannot know the exact value of  $R$  for each project. As a result, they cannot do price discrimination among firms with different projects. Second, it is probably true in reality that commercial banks do know *some* information about the profitability of each project it funds. This correspond to the assumption of our model that the incumbent bank knows whether the profitability of each project is above a certain threshold which is the minimum requirement for firms to receive loans.<sup>7</sup> We include the cost to acquire such a piece of information into the cost of raising money from the public which is normalized to one. Finally, it is probably true in reality that it is *increasingly* costly for commercial banks to acquire finer and finer information on the projects they fund. Therefore, they choose to acquire some *rough* information about the funded projects. In the model, we assume that the commercial bank only acquires one piece of information in equilibrium. That is whether the profitability of the project is above a certain threshold for providing funding. We abstract from complicated analysis of the optimal information acquisition decision in the paper, since our focus is on the interaction between the political contribution and equilibrium government policy.<sup>8</sup>

---

<sup>7</sup>This threshold is denoted as  $R(r)$  in subsequent analysis.

<sup>8</sup>We can rationalize the assumption adopted in the paper by assuming that the cost to acquire multiple pieces of information (e.g., two interest rates and two thresholds) is too costly for the bank to pay.

For expositional purposes, we make several simplifying assumptions. First, we assume that there is only one commercial bank in each state, and therefore that bank is a monopolist when the state government chooses not to liberalize the banking sector. Second, when the branching deregulation is implemented, multiple banks enter the banking sector which results in Bertrand competition. As a result, the gross interest rate in equilibrium is pushed down to one, which is assumed to be the cost of raising funds from the public. Third, as our focus is political influence from the banking sector within the state, we abstract from the possibility of political contributions from firms and commercial banks outside that state. Fourth, as it is a static model, we do not want to create a dynamic linkage between periods to allow the commercial bank to hedge their contribution against the economic environment. This assumption is reasonable, as most politicians are short-lived players in the political arena, or at least they are very much unsure whether they can stay in power for long enough time. Fifth, we assume that the bank's unit cost of raising funds from the public is constant and equals one. Finally, we assume that each state can only move its policy from branching regulation to branching deregulation. In other words, the state government can't regulate a banking sector that has already been freed from the branching regulation. This is a realistic assumption, as the cost of regulating an industry that has already been freed from regulation is prohibitively high due to the pressure posed by other states or countries.

## 2.2.2 Financial Friction and Credit Rationing

Firms have to obtain external finance in order to invest in projects, and some of them end up receiving no external finance due to their low profitability. More specifically, each firm with the profitability level of  $R$  needs to have  $I$  amount of money to fund the project. As long as the firm invests  $I$ , it will obtain the revenue of  $RI$  for sure after the project is completed. Each firm's asset is  $A$  which is assumed to be smaller than  $I$ . Accordingly, every firm needs to raise  $I - A$  amount of money to make the investment. The parameter of  $\underline{R}$  is assumed to be bigger than or equal to one, which means in the first best world, all firms should be funded. There is one type of financial friction, however, which generates the credit rationing. Following the convention in the macroeconomics literature, we assume that when the project is completed, the firm's owner can run away with  $(1 - \theta)RI$  as his revenue and does not pay the loan it receives from the bank. Therefore, the bank is willing to lend money to a firm with the profitability level  $R$  if the following condition is satisfied:

$$(1 - \theta)RI \leq RI - (I - A)(1 + r), \quad (2.1)$$

where  $r$  is net interest rate the bank charges on each firm. The left hand side of the above inequality is the net payoff of the firm owner when he runs away, while the right hand side is his net payoff when the firm pays bank the loan. It is straightforward to observe that a firm with a bigger  $R$  is more likely to obtain the credit, since a bigger  $R$  means a larger loss for the owner when he runs away. The cutoff on the profitability level  $R$  above which the firm obtains the external finance is determined



by

$$R(r) = \left(1 - \frac{A}{I}\right) \frac{(1+r)}{\theta}. \quad (2.2)$$

In order to have interesting results, we assume that  $\underline{R} < (1 - \frac{A}{I})\frac{1}{\theta}$ . Thus, even if the gross interest rate equals the bank's unit cost of raising money from the public, the bank still does not have the incentive to fund all the projects. Therefore, we can not be in the first best case. We also assume that  $\bar{R} > (1 - \frac{A}{I})\frac{1}{\theta}$ . This means at least the most profitable firm could get the external finance to undertake its project, when the interest rate reflects the cost of raising funds from the public. It is easy to see that a higher interest rate  $r$  means a higher cutoff for firms to obtain external finance. Consequently, in the second best case in which the social welfare is maximized under the financial friction (i.e., the firm owner's running away), the net interest rate should be zero.

We consider the case under branching regulation first. In this scenario, the commercial bank is a monopolist in the credit market and consequently can implement monopoly pricing. As the commercial bank's unit cost of raising money from the public is assumed to be one, its objective function is

$$\max_r r(I - A)[1 - G(R(r))],$$

where  $G(\cdot)$  is the cumulative density function of  $R$ . Simple calculation shows that the first order condition is

$$[1 - G(R(r^*))] - r^*g(R(r^*))\left(1 - \frac{A}{I}\right)\frac{1}{\theta} = 0,$$

where  $g(\cdot)$  is the probability density function of  $R$ . In order to have some insights from the model, we adopt a specific distribution function of  $R$ : a uniform distribution on  $[\underline{R}, \bar{R}]$ .<sup>9</sup> The equilibrium interest rate can be calculated as

$$r^*(\theta, \bar{R}) = \frac{\bar{R} - (1 - \frac{A}{I})\frac{1}{\theta}}{2(1 - \frac{A}{I})\frac{1}{\theta}}. \quad (2.3)$$

The basic tradeoff the commercial bank faces is between having more loans which is consistent with a lower interest rate and making more profit from each loan which is consistent with a higher interest rate. As the equilibrium interest rate is strictly positive, we are not even in the second best case now.<sup>10</sup> It is straightforward to observe that  $r^*(\theta, \bar{R})$  increases in  $\bar{R}$ . The intuition is that a higher profitability level on average makes the commercial bank care more about a higher profit per loan. Similarly, the reason why  $r^*(\theta, \bar{R})$  increases in  $\theta$  is that a shrinking pool of firms (due to a smaller  $\theta$ ) that can get access to external finance makes the commercial bank tilt its consideration in favor of having more loans. Substituting the equilibrium interest rate in Equation (2.3) into Equation (2.2), we can derive the cutoff for external financing under branching regulation as

$$R^* \equiv R(r^*) = \frac{\bar{R} + (1 - \frac{A}{I})\frac{1}{\theta}}{2}. \quad (2.4)$$

The analysis in the case without branching regulation is straightforward. The entry of the second commercial bank into the market drives the net equilibrium

---

<sup>9</sup>The second order condition for the above optimization problem requires that  $g(\cdot)$  is a non-decreasing function. The uniform distribution satisfies this property.

<sup>10</sup>Remember we assume that  $\bar{R} > (1 - \frac{A}{I})\frac{1}{\theta}$ .

interest rate to zero (i.e., the cost of raising money from the public), which we denote it by  $r_0$ . Accordingly, the cutoff for receiving external financing in this case is

$$R_0 = \left(1 - \frac{A}{I}\right) \frac{1}{\theta}.$$

Obviously, we have  $r^* > r_0 \equiv 0$  and  $R^* < R_0$ .

### 2.2.3 Political Contribution and Branching Deregulation

We discuss the equilibrium government policy in this subsection. Following Grossman and Helpman (1994), we can derive the government's objective function as

$$(1 - \lambda)\Pi(i) + \lambda C(i), \tag{2.5}$$

where  $\Pi(i)$  is the sum of the commercial bank's profit and the profit of firms (i.e., the social welfare);  $C(i)$  is the political contribution based on the policy choice  $i$ . In the above equation,  $i = 1$  means branching regulation and  $i = 0$  means branching deregulation.  $\lambda$  is the weight the government puts on contribution, and a lower  $\lambda$  indicates a more benevolent government.

As the government can always abandon the branching regulation policy, all of the commercial bank's profit should be extracted by the government.<sup>11</sup> Therefore,

---

<sup>11</sup>To make the model more realistic, one can add a Nash bargaining between the government and the incumbent bank. In that case, *a fraction* of the bank's profit accrues to the government. However, as long as the bargaining power is fixed, the weights in Equations (2.5) and (2.6) are unaffected by the government's policy choice. Therefore, this alternative specification does not affect qualitative results of the model.

we can rewrite the government's objective function as

$$(1 - \lambda)\Pi(i) + \lambda\Pi_B(i), \quad (2.6)$$

where  $\Pi_B(i)$  is the profit of the commercial bank within that state. In the case of branching deregulation, the firms' total profit and the commercial bank's profit are

$$\Pi(0) = \int_{R_0}^{\bar{R}} \frac{(R-1)I}{\bar{R}-\underline{R}} dR = \frac{I}{\bar{R}-\underline{R}} \left[ \left( \frac{\bar{R}^2}{2} - \bar{R} \right) - \left( \frac{R_0^2}{2} - R_0 \right) \right] \quad (2.7)$$

and

$$\Pi_B(0) = 0. \quad (2.8)$$

In the case of branching regulation, the firms' total profit and the commercial bank's profit are

$$\Pi(1) = \int_{R^*}^{\bar{R}} \frac{(R-1)I}{\bar{R}-\underline{R}} dR = \frac{I}{\bar{R}-\underline{R}} \left[ \left( \frac{\bar{R}^2}{2} - \bar{R} \right) - \left( \frac{R^{*2}}{2} - R^* \right) \right] \quad (2.9)$$

and

$$\Pi_B(1) = \frac{I}{\theta(\bar{R}-\underline{R})} \left[ \theta\bar{R} - \left( 1 - \frac{A}{I} \right) \right]^2. \quad (2.10)$$

After substituting Equations (2.7)-(2.10) into Equation (2.6), we can compare the payoffs of the government in the two cases. Branching deregulation is preferred to regulation if and only if

$$\frac{1-\lambda}{\lambda} \left[ \left( \frac{R^{*2}}{2} - R^* \right) - \left( \frac{R_0^2}{2} - R_0 \right) \right] \geq \frac{1}{\theta} \left[ \theta\bar{R} - \left( 1 - \frac{A}{I} \right) \right]^2, \quad (2.11)$$

or

$$L(\theta, \bar{R}) \equiv \frac{\theta(1-\lambda)\left[\left(\frac{R^{*2}}{2} - R^*\right) - \left(\frac{R_0^2}{2} - R_0\right)\right]}{\lambda\left[\theta\bar{R} - \left(1 - \frac{A}{I}\right)\right]^2} = \frac{(1-\lambda)(R^* + R_0 - 2)}{8\theta\lambda(R^* - R_0)} \geq 1. \quad (2.12)$$

In other words, the government prefers liberalizing the banking sector, if the increase in firms' total profit exceeds the decrease in the commercial bank's profit substantially after branching deregulation.

Parameter  $\theta$  is closely related to a country (or a state)'s institutional quality and has implications for banking deregulation. It can be proved that  $\frac{\partial L(\theta, \bar{R})}{\partial \theta} < 0$ . In other words, a country with a better institution (i.e., a higher  $\theta$ ) tends to regulate the banking sector, as a better institution (e.g., strong protection for creditors) gives the banking sector more opportunities to extract profit from firms in the product market. As the profitability levels of all firms do not change, the commercial bank's profit increases faster than the firms' total profit when the quality of institution is improved. This result has two implications. First, we have to differentiate between the quality of the institution and a specific policy implemented in the banking sector when we do empirical work. As the model shows, a better institution is more likely to be associated with a regulated banking sector, which *seems* to indicate that the institutional quality is low. Second, it may be not surprising to see some advanced countries like Japan to have a highly regulated banking sector, even if these countries have very good financial institutions to prevent borrowers from escaping.

From Freeman (2002), we know that banking deregulation usually happens in recessions. This model can give a rationale for this fact. In the theory, a recession

means a decrease in the mean of the profitability distribution without any change in the variance of it. For the case of a uniform distribution, this leads to the same degree of decrease in both  $\underline{R}$  and  $\bar{R}$ ]. Then, it is easy to observe that  $\frac{\partial L(\theta, \bar{R})}{\partial \bar{R}} < 0$ . In other words, the government is more likely to deregulate the banking sector when the state's economy is in recession. The key observation for this result is that in the case of branching regulation, the share of the commercial bank's profit in the total social welfare increases when the economy moves from a recession to a boom, since

$$\begin{aligned} & \frac{\Pi_B(1)}{\Pi(1)} \\ = & \frac{r^*(I - A)[1 - G(R^*)]}{\int_{R^*}^{\bar{R}} \frac{(R-1)I}{R-\underline{R}} dR} \\ = & \frac{8\theta[\bar{R} - R_0]}{3\bar{R} + R_0 - 4}, \end{aligned}$$

increases in  $\bar{R}$ .<sup>12</sup> This implies that the importance of the commercial bank's profit (relative to firms' profit) for the government increases when the economy moves from a recession to a boom. Consequently, the political contribution which equals the commercial bank's profit matters more for the government. This incentivizes the government to regulate the banking sector in order to receive the political contribution.

---

<sup>12</sup>Remember that  $R_0 > \underline{R} \geq 1$ .

## 2.2.4 Implications for the Middle Income Trap

The above result has implications for the middle income trap which is the most important contribution of our paper. The existing explanation for the middle income trap is mainly from the technological perspective. Namely, underdeveloped economies can use imitation and technological transfers to boost their economic growth at the early stage of the development, since they are far away from the world technological frontier. However, when their economic development achieves a certain level, developing countries have to innovate by themselves. This may be hard for them to do, since they are used to imitating and receiving technology transfers from developed countries. This eventually causes the middle income trap. Our story in this paper offers an alternative explanation for the middle income trap from the perspective of political economy. Political incumbents have a greater incentive to distort the resource allocation and receive rents only when the economic development achieves a certain level. The reason is simple, since they can extract few rents when the economy is poor. However, the greater incentive to distort the resource allocation in the market prevents the economy from further development, which creates the middle income trap in our story.

The alternative explanation for the middle income trap receives support from existing evidence and deserves more attention in the future research. Economic historians such as Buchanan, Tullock and North found that institutions play a big role in determining the long-run economic development (Buchanan and Tullock, 1963; North, 1981 and 1990). Moreover, North (1990) argued that institutions themselves are endogenously determined by economic agents such as the government and the

special interest groups. Our paper argues that the above two insights are crucial for us to understand the middle income trap phenomenon as well. In particular, the simple model presented in this paper points out *an exact channel* through which political influence leads to the middle income trap. That is, the government cares both social welfare and political contributions. Only when the economy achieves a certain level of development which enables the special interest group to contribute enough money to the government, does the government have a big enough incentive to design a distorted policy and extract rents from the economy. As a result, countries that do not restrict the political contribution to a low level are less likely to get out of the middle income trap. This helps explain the different development trajectory of economies with similar technological and demographic starting points (e.g., Malaysia and Indonesia vs Hong Kong and Singapore). Using cross-country data on economic performance and institutions, Aiyar *et al.* (2013) found that countries that reduce government involvement in the economy and deregulate their credit markets are more likely to get out of the middle income trap. This empirical finding directly support the argument advocated in this paper. Moreover, the recent experience of China should draw our attention to the role of institutions in determining whether a country can get out of the middle income trap. I leave this potentially fruitful research agenda for future research.



## 2.3 Empirical Evidence

There are several theoretical predictions we want to test in this section. In particular, we are interested in testing the following four predictions, as they are all the key predictions of the model. First, are recessions correlated with branching deregulation? Second, is banks' profit more volatile than the state's GDP? Third, does small banks' profit (i.e., unit banks' profit) decrease after branching deregulation? Finally, do commercial banks' political contribution moves one-to-one with their profit? We investigate these questions in the following subsections.

### 2.3.1 Deregulation and GDP

In the model, GDP of a state increases after branching deregulation's happening in the banking sector. This phenomenon is well documented in several influential papers (e.g., Jayaratne and Strahan (1996), Cetorelli and Strahan (2006)). Another prediction of our model that branching deregulation is more likely to happen in recessions is consistent with Freeman (2002). Figure 1 from Freeman (2002) clearly shows this pattern.

### 2.3.2 Banks' Profit and GDP

This subsection verifies the result that profit of the commercial banking sector positively co-moves with GDP at the state level, and is more volatile than GDP. We regress the growth rate of deflated net income of all commercial banks (*grttni*) within a state on real GDP (*grrgdp*) at the state level. The result is reported in

Table 2.1. The coefficient in front of *grrgdp* is positive and statistically significant. Furthermore, the elasticity of the deflated net income of all commercial banks within the state with respect to real GDP at the state level is almost six. This implies that the banking sector's profit at the state level is much more volatile than real GDP at the state level. In short, the evidence presented in Table 2.1 support one key result of our model that the banking sector's profit is more volatile than GDP.

Table 2.1: Growth Rate of Bank Profit and GDP at the State-Level

	(1)	(2)	(3)
	<i>GrowBank</i>	<i>GrowBank</i>	<i>GrowBank</i>
<i>GrowGDP</i>	5.133** (2.58)	6.343** (2.23)	5.378** (2.54)
_cons	-0.127*** (-3.31)	-0.159 (-1.27)	0.0765 (0.42)
State F.E.	Y	N	Y
Year F.E.	N	Y	Y
<i>N</i>	918	918	918
<i>R</i> <sup>2</sup>	0.053	0.019	0.070

*GrowBank*: annual growth rate of deflated profit of banks.

*GrowGDP*: annual growth rate of real GDP.

Years covered: 1976-1994.

*t* statistics in parentheses.

Standard errors are clustered at the state level.

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

### 2.3.3 Deregulation, Net Interest Margin and Banks' Profit

The assumption of Bertrand competition (after branching deregulation) is only for the illustrative purpose. The essence of this assumption is that intensified compe-

tition after the branching deregulation drives down the profit of commercial banks within the state. We provide empirical evidence for this result in this subsection. We use bank-level data from Bank Regulatory data set of WRDS which comes from FDIC call reports to test this prediction. Data period for the analysis is from 1976 to 1995, during which most states removed their bank branching restrictions. Since data in call reports are year-to-date, we only utilize data reported in December. In reality, unit banks, which do not set up branches elsewhere, are the parallel of the incumbent bank in our model. Although there is a variable (item 9421) in the data set identifying the charter type of each bank (e.g., state or federal), there is no such a variable indicating how many branches each bank has. Because of this limitation, we apply an identifying assumption that unit banks are smaller in asset size to the following analysis.

In the regressions, we explore the relationship between several indicators for banks' profit and the branching deregulation. The first dependent variable we look at is the Net Interest Margin (NIM), defined as the ratio of net interest income (item 4010 minus item 4170) to total loan (item 1400). We focus on banks with state-level charter and drop observations with  $|NIM| > 1$ . We regress  $NIM$  on  $DEG$ ,  $\log(asset)$ , and  $DEG * \log(asset)$ , where  $DEG_{s,t}$  is a dummy variable. It equals one, if the deregulation in the state where bank  $s$  is headquartered happens no later than year  $t$ . Variable  $\log(asset)_{s,t}$  is the natural log of total assets of bank  $s$  in year  $t$ . In specification one, we control for year and state fixed effects; in specification two, we control for year and bank fixed effects. The results are shown in Table 2.2. Next, we replace  $NIM$  with  $\log(profit)$ , the natural log of profit from borrowing

and lending, defined as  $\log(\text{item 4010 minus item 4170})$ . Similarly, in specification three, we control for year and state fixed effects; in specification four, we control for year and bank fixed effects. The results are also shown in Table 2.2.

The estimation results lend support to our theoretical predictions. In all specifications, the coefficient in front of  $DEG$  has a negative sign and is significant at 1% level. Furthermore, the estimated coefficient is economically significant as well. This indicates that competition in the banking industry due to branching deregulation becomes more severe after deregulation. The coefficient in front of  $\log(asset)$  has a positive sign in all specification, but is not always statistically significant. The coefficient in front of  $DEG * \log(asset)$ , which is related to one key prediction of our model, always has a positive sign, and is significant at 1% level. This shows that deregulation affected banks with different sizes differently. On the one hand, small banks, which are close to the incumbent bank in our model,<sup>13</sup> were adversely affected by the deregulation in terms of NIM and profit. On the other hand, banks with large-scale assets actually *gained* from the branching deregulation. Note that the timing of the branching deregulation decision's being made was correlated between states. Therefore, for a big bank that can set up branches outside its own state, the loss in profit earned within the state (due to the branching deregulation within the state) might be smaller than the gain in profit earned from other states that deregulated their banking sectors around the same time as its own state. This explains why big banks actually gained in profit when their own states deregulated the banking sectors. In some sense, this empirical pattern for commercial banks

---

<sup>13</sup>This is because they are less likely to earn profit from multiple states or regions.

echoes the composition effect of trade liberalization on firm size documented extensively in international trade literature.<sup>14</sup> As a robustness check, we add GDP growth rate of the state where the bank is headquartered into the regression, and the results are in columns five to eight in Table 2.2 . Although GDP growth rate has a negative sign in all specifications and is significant except for specification seven, it does not change the results obtained in the previous four regressions qualitatively.

### 2.3.4 Banks' Profit and Campaign Contribution

In this subsection, we investigate the relationship between banks' profit and their campaign contribution. As both Grossman and Helpman (1994) and our theory predict, the elasticity of the banking sector's political contributions with respect to its profit should be one.<sup>15</sup> We show evidence for this theoretical result in this subsection. As the branching deregulation decision is made at the state level, ideally we want to use data on state-level campaign contributions. However, to our best knowledge, such data before year 2000 are unavailable. Since FEC digitalized *federal* level campaign contribution data since 1979, we make use of that to estimate the elasticity of the campaign contributions from the commercial banking sector of each state with respect to total net income of that sector. Data period for this analysis is from 1979 to 1992. Net income data come from the FDIC data set used in the

---

<sup>14</sup>In the trade literature on heterogeneous firms, a common result is that exporting and big firms gain market shares, while non-exporting and small firms lose market shares after bilateral trade liberalization. See Melitz (2003) for more details etc.

<sup>15</sup>This is true, even if we allow for a Nash bargaining between the government and the banking sector.



previous subsection (item 4340). Both the political contribution and the net income are deflated.

Due to data limitation, we only include campaign contributions made by PACs which account for roughly 40% of the total contribution into the regressions. The rest is made by individuals, which we are not able to identify whether it is on behalf of commercial banks. It is likely that contributions made by PACs well represent the elasticity we are interested in, and we can infer the change in the total contribution from this fraction of it. PACs' related banks are identified using the catalog from Center for Responsive Politics,<sup>16</sup> and it is assumed that the state where an PAC is located is the same as where its related bank is.

The regression results are supportive of our model. We regress  $\log(rcont)_{s,t}$  (the natural log of total real campaign contributions made by commercial banks' PACs of state  $s$  in year  $t$ ) on  $\log(rttni)_{s,t}$  (the natural log of total net income of commercial banks of state  $s$  in year  $t$ ). We do not control for state and year fixed effects in the regression, as the theory only predicts a one-to-one relationship between bank profit and political contributions irrespective of other factors. The estimation results for specifications with and without the time fixed effect are reported in Table 2.3. As a robustness check, we exclude DC in specifications three and four, since the data pattern of the nation's capital may be different. The elasticity estimated is significant at 1% level, and is very close to one. The null hypothesis that the estimated coefficient equals one can't be rejected. This lends support to our theoretical prediction.

---

<sup>16</sup>This is an independent non-profit organization, whose data are also used by other papers such as Kroszner and Stratmann (1998).

Table 2.3: Political Contributions and Bank Profit

	(1)	(2)	(3)	(4)
	$\ln(\text{contribution})$	$\ln(\text{contribution})$	$\ln(\text{contribution})$	$\ln(\text{contribution})$
$\ln(\text{net income})$	0.913*** (5.69)	0.932*** (5.86)	0.982*** (6.74)	0.998*** (6.91)
_cons	4.426*** (4.04)	3.827** (3.41)	3.895*** (4.04)	3.259** (3.36)
Time F.E.	No	Yes	No	Yes
D.C.	included	included	excluded	excluded
$N$	309	309	304	304
$R^2$	0.300	0.328	0.374	0.402

Years covered: 1980, 1982, 1984, 1986, 1988, 1990, 1992.

Standard errors are clustered at the state level.

$t$  statistics in parentheses

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

## 2.4 Conclusion

We set up a simple model to rationalize the fact that the government has a greater incentive to deregulate the banking sector when the economy is in boom. This is because the rent that can be extracted from the banking sector to the government (under regulation) increases disproportionately more compared with the loss in social welfare due to regulation, when the economy moves from a recession to a boom. We can use data on political contributions and banks' profit to test three theoretical predictions of the model and found that all of them gain support from the data. In particular, we find that the impact of branching deregulation on banks' profit is heterogeneous with a bigger negative effect on small banks. In addition, using bank-level contribution data, we find that the banking sector's total political



contribution moves one-to-one with its total profit, which is a key implication of Grossman-Helpman type political economy models.

The economic story articulated in this paper has implications for the middle income trap. Especially, we offer an explanation for the middle income trap from the perspective of political economy. That is, only when the economy achieves a certain level of development, does the government have incentives to design distorted policies and extract rents from the economy. These distorted policies reduce social welfare and hinder further development of the economy which leads to the middle income trap. We are confident that this new explanation should receive more attention in the near future, especially given what has happened in China in the past decade. Further empirical analysis is needed to show the validity of the political economy story offered in this paper for explaining the middle income trap.

Needless to say, there is still ample work to be done. First, in a more realistic scenario, we can allow the unit bank as a monopolist to offer *a menu* of interest rates to the borrowers subject to an information acquisition cost it has to pay. Ideally, we want to show that offering a uniform interest rate is the best strategy for the bank to take. Second, a dynamic model might be useful to explain why the government can't move from deregulation to regulation when the economy enters into a boom again. Finally, better data are needed to show the robustness of the empirical results we have obtained above.

## 2.5 References

1. Aiyar, S., Duval, R., Puy, D., Wu, Y., and Zhang, L., 2013. Growth slowdowns and the middle-income trap. IMF Working Paper: WP/13/71.
2. Bloom, D., and Williamson, J., 1998. Demographic transitions and economic miracles in emerging Asia. *World Bank Economic Review*, 12 (3): 419-455.
3. Bloom, D., and Canning, D., 2004. Global demographic change: Dimensions and economic significance, in *Global Demographic Change: Economic Impacts and Policy Challenges*, proceedings of a symposium, sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 26–28, pp. 9-56.
4. Brandt, L., Van Biesebroeck, J., and Zhang, Y., 2012. Creative accounting or creative destruction: Firm level productivity growth in Chinese manufacturing. *Journal of Development Economics*, 97 (2), 339-351.
5. Brandt, L., Tombe, T., and Zhu, X., 2013. Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics*, 16 (1), 39-58.
6. Buchanan, J., and Tullock, G., 1962. *The calculus of consent: Logical foundations of constitutional democracy* (Ann Arbor: University of Michigan Press).
7. De Long, J., and Shleifer A., 1993. Princes and merchants: European city growth before the industrial revolution. *Journal of Law and Economics*, Vol. 36 (2), 671-702.

8. Cetorelli, N., and Strahan, P., 2006. Finance as a barrier to entry: Bank competition and industry structure in U.S. local markets. *Journal of Finance*, 61 (1), 437-461.
9. Freeman, D., 2002. Did state bank branching deregulation produce large growth effects? *Economics Letters*, 75 (3), 383-389.
10. Grossman, G., and Helpman E., 1994. Protection for sale. *American Economic Review*, 84 (4), 833-850.
11. Goldberg, P., and Maggi G., 1999. Protection for Sale: An empirical investigation. *American Economic Review*, 89 (5), 1135-1155.
12. Huang, Y., 2008. *Capitalism with Chinese Characteristics: Entrepreneurship and the State*. Cambridge University Press.
13. Jayaratne, J., Strahan, P., 1996. The finance-growth nexus: Evidence from bank branch deregulation. *Quarterly Journal of Economics*, 111 (3), 639-670.
14. Khandelwal, A., Schott, P., and Wei, S., 2013. "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," *American Economic Review* 103: 2169-2195.
15. Kroszner, R., 2001. The motivations behind banking reform. *Regulation*, 24 (2), 36-41.
16. Kroszner, R., and Strahan, P., 1999. What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *Quarterly Journal of Economics*, 114 (4), 1437-1467.

17. Kroszner, R., and Stratmann, T., 1998. Interest-group competition and the organization of congress: Theory and evidence from financial services' political action committees. *American Economic Review*, 88 (5), 163-187.
18. Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71 (6), 1695-1725.
19. North, D., 1981. *Structure and change in economic history*. (New York: Norton Company).
20. North, D., 1990. *Institutions, institutional Change, and economic performance*. (Cambridge: Cambridge University Press).
21. Rajan, R, and Zingales, L., 2003. The great reversals: The politics of financial development in the 20th century. *Journal of Financial Economics*, 69 (1), 5-50.
22. White, E., 2000. Banking and finance in the twentieth century, in Stanley L. Engerman and Robert E. Gallman, eds., *The Cambridge Economic History of the United States: The Twentieth Century Volume III* (Cambridge: Cambridge University Press).

# Chapter 3

## Contract Design under Routing Auditing

### 3.1 Introduction

In financial markets, often times an entrepreneur has better information about the profitability of his projects than his external investors. And usually an external investor can prevent potential fraudulent misrepresentation of profits by the entrepreneur only by costly auditing. Pioneered by Townsend (1979) and Gale and Hellwig (1985), the costly state verification literature formally models this scenario, and probes into how to mitigate the moral hazard problem and reduce the socially wasteful auditing cost involved by optimally designing the contract between the entrepreneur and the investor. In such models, the investor decides whether or how to audit the firm after observing the financial report by the entrepreneur, and the

optimal contract reduces the investor's incentive to audit and thus saves the cost involved by making the transfer to him least sensitive to the realization of the firm's profit and to the report of the entrepreneur. As a result, a standard debt contract typically emerges as the optimal contract.<sup>1</sup>

The key driving force of the results of these models is the state-contingency of the investor's auditing decision. However, this is not a good approximation of the reality under many circumstances. For example, there are legal requirements mandating firms to be routinely audited, regardless of their profitability; An investor may simultaneously invest in many firms and have limited attention to keep track of the timing and contents of their financial reports and make his auditing decision accordingly; It may be hard for different investors of a same company to coordinate their auditing decisions under various contingencies. What if the investor's auditing decision has to be non-contingent on the report of the entrepreneur (hereafter, routing auditing)? This is the question this chapter tries to tackle.

Specifically, to highlight the role played by the state-contingency of auditing decisions, we consider the contract design implication of routing auditing in a model resembling Gale and Hellwig (1985) otherwise. In the model, an entrepreneur wants to cash out an asset that generates a random positive cash flow in the future, the amount of which can only be observed by himself costlessly upon its realization. She first designs and offers to an investor a take-it-or-leave contract which specifies the price the investor pays right away, and future transfer to the investor contingent on the entrepreneur's profit report and on the realization of cash flow if also observed

---

<sup>1</sup>Some additional examples include Mookherjee and Png (1989), Winton (1995), and Krasa and Villamil (2000).

by the investor in the auditing. If the investor accepts the contract, she chooses an intensity of auditing *before* the entrepreneur makes her profit report. Then, the entrepreneur makes her profit report upon observing the realization of cash flow, and the transfer is made according to the contract.

The main result of this chapter is that, in this otherwise standard Gale and Hellwig (1985) model, an equity contract turns out to be optimal. This contrasts the original Gale and Hellwig (1985) model, which yields debt as the optimal contract. Why is this the case? Note that there is a time-consistency issue with the entrepreneur. Ex post, she tries her best to minimize her transfer to the investor. But ex ante, she wants to maximize the transfer in order to enjoy as much gain from trade as possible. When designing the contract ex ante, she understands that the punishment for lying when caught is her only commitment device: Whatever auditing intensity the investor will choose, letting him take over all the cash flow if she is caught lying maximizes the transfer enforceable ex post and credible ex ante. Here, the non-contingency of the auditing technology plays the key role: Since the auditing intensity  $p$  is constant over all contingencies, the enforceable transfer for each realization of cash flow  $x$  is  $px$ , which takes the form of an equity. When designing the contract, it is irrational to make the de jure transfer less than the enforceable transfer  $px$ , because that incentivizes the entrepreneur to tell the truth at period 1 and reduces the transfer. And the easiest way to make the de jure transfer greater than the enforceable transfer for each realization of cash flow  $x$  is to make it also an equity, with the proportion to the investor greater than the expected auditing inten-

sity. This conclusion is robust, in the sense that it does not rely on the distribution of cash flow or on the functional form of the auditing cost function.

The contribution of this finding can be viewed in three different ways. First, it complements the existing literature of costly state verification by considering the other extreme scenario of verification technology, i.e., verification decision cannot be state-contingent, and thus highlights the key role of state contingency of verification decision in shaping the design of optimal contracts. Second, it provides advice to practitioners who face the task of designing financing contracts when auditing is not able to be carried out contingent on the self-report of the entrepreneur. Last, and most importantly, it provides a novel and simple answer to a longstanding problem in contract theory: Why are linear contracts so common in practice, while textbook models often predict more complicated functional forms? Moreover, according to the classic pecking order theory (e.g. Myers and Majluf, 1984), equity is more costly than other ways of external financing, e.g. debt and convertible bonds. An early explanation was offered by Holmstrom and Milgrom (1987). In their model, the agent controls the mean of the profitability of the project in a dynamic setup, which is not perfectly observable by the principal. Although the principal can make payments depend on the entire path of motion, the optimal contract is simply a linear function of the end point, due to the stationary structure of the model implied by CARA utility. One strand of existing literature invokes the robustness value of linear contracts in the context of moral hazard. When the principal faces uncertainty about the technology of the agent, linearity of the contract, which fixes the ratio of the agent's payoff to the principal's, serves as the tool for the principal to turn



his assurance about the agent’s payoff into a guarantee for himself.<sup>2</sup> Axelson (2007) instead considers the context of adverse selection. He shows that when investors rather than the manager have private information about the firm or project, it is often optimal to issue information-sensitive securities such as equities. This chapter complements these works by invoking the optimality of equity contract in a different and realistic context—when the agent may lie about the actual profitability of the project and the principal cannot audit the agent on a report-contingent basis. In this case, the auditing intensity is by construction constant across different states, and it is optimal for the entrepreneur to maximize the punishment when designing the contract to reduce her temptation to lie ex post. Thus, the de facto transfer enforceable by such auditing technology takes the form of equity, which is the key driving force of the result.

The rest of this chapter is organized as follows. Section 2 introduces the model setup. Section 3 proves the result. And Section 4 concludes.

## 3.2 The Model

We consider a two-period game with two players: an entrepreneur and an investor. The entrepreneur is endowed with an asset at period 0, which generates a random cash flow  $x \in [0, \bar{x}]$ <sup>3</sup> at period 1. The (potential) investor holds consumption goods (money) at period 0. Following the convention of the security design literature, we

---

<sup>2</sup>For example, see Diamond (1998), Chassang (2014), and Carroll (2015).

<sup>3</sup> $\bar{x}$  can be infinity. Note that it is without loss of generality to assume the lower bound of the support of  $x$  to be 0, because if it is some  $\underline{x} > 0$ ,  $\underline{x}$  will be sold as risk-free debt at price  $\delta_b \underline{x}$ , and the support of the remaining cash flow will then have lower bound of 0.

assume that both players are risk neutral. Specifically, player  $j$ 's utility is given by  $u_j = c_{j0} + \delta_j c_{j1}$ , where  $c_{jt}$  denotes player  $j$ 's consumption at period  $t$ , and  $\delta_j$  is his subjective discount factor,  $j \in \{e, i\}$  ( $\{e, i\}$  stands for  $\{\text{entrepreneur, investor}\}$ ). We assume  $\delta_i > \delta_e$ , i.e. the entrepreneur has a better investment opportunity than the investor.<sup>4</sup> This assumption creates the trading demand: Both players may benefit from transferring some goods to the entrepreneur at date 0 and compensating the investor with repayment backed by the random cash flow  $x$  at period 1.

As in Gale and Hellwig (1985), at period 0, there is no information asymmetry: both players have the same knowledge of the distribution of  $x$ . The only restriction on the distribution of  $x$  is that it has a well-defined mean  $E[x]$ . The entrepreneur offers a take-it-or-leave contract  $s = (q, r(\cdot), R(\cdot, \cdot))$  to the investor, who decides whether to accept it.  $q$  is the price paid by the investor to the entrepreneur at period 0.  $r(\hat{x})$  is the period 1 transfer from the entrepreneur to the investor if the entrepreneur then reports the cash flow to be  $\hat{x}$  and the investor is not able to observe the true realization of  $x$ . If the investor instead observes the true realization of  $x$  in period 1, the entrepreneur pays the investor  $R(x, \hat{x})$  then. Feasibility requires  $r(\hat{x}) \in [0, \hat{x}]$  and  $R(x, \hat{x}) \in [0, x]$ . If the investor rejects the contract, the game ends here, and the entrepreneur's and the investor's payoffs are  $\delta_e E[x]$  and 0 respectively. Since our focus is on optimal contract design, we restrict the parameter values to be such that there is at least a contract acceptable by the investor in equilibrium.

There are three sub-periods at period 1. In the first sub-period, the investor chooses the intensity of his auditing on the realization of the entrepreneur's cash flow

---

<sup>4</sup>Another interpretation is that the seller has a higher carrying cost than the buyer, as in Hennessy (2012).

$x$ . The intensity is modeled as the probability  $p$  of observing the true realization of  $x$ . With the complementary probability  $1 - p$ , the investor observes nothing from auditing. There is a cost  $c(p)$  incurred in the auditing, no matter whether the truth is observed.  $c(0) = 0$ ,  $c' > 0$ ,  $c'' > 0$ , and  $c(1)$  is sufficiently large, to rule out the uninteresting possibility that the investor chooses to observe the truth for sure.

In the second sub-period, the entrepreneur observes the realization of the random cash flow  $x$  costlessly, and also the investor's choice of auditing intensity  $p$ , but not whether the investor actually sees the true  $x$ . Based on that, she chooses her report  $\hat{x}$  to the investor of cash flow realization, which may or may not be truthful.

In the last sub-period, the investor observes the true realization of  $x$  with the probability  $p$  chosen earlier. And the transfer from the entrepreneur to the investor is made according to the contract specified in period 0.

We look for subgame perfect equilibria, in which: 1) The contract designed in period 0 maximizes the entrepreneur's expected payoff; 2) In period 1, the investor's choice of auditing intensity  $p$  maximizes his payoff; and 3) In period 1, the entrepreneur's reporting strategy maximizes her payoff.

Note that other than the auditing technology, this model resembles Gale and Hellwig (1985). By having the investor making his auditing decision before the entrepreneur makes her report, the auditing decision is made non-contingent on the realization of cash flow  $x$  and on the entrepreneur's report  $\hat{x}$ . This captures the key characteristic of routine auditing mentioned in the introduction.

### 3.3 Equity as Optimal Contract

The purpose of this subsection is to prove that equity is an optimal contract in this model. It takes two steps to establish the optimality of an equity contract. First, we prove that for all the contracts the entrepreneur can choose in period 0, there is a common upper bound for her expected payoff. Second, we propose an equity contract and show that it achieves the upper bound in the previous step.

We solve the model by backward induction. At period 1, given the investor's auditing intensity  $p$  and the true realization of cash flow  $x$ , the entrepreneur chooses her report  $\hat{x}$  to minimize the expected transfer from her to the investor:

$$\min_{\hat{x}} pR(x, \hat{x}) + (1 - p)r(\hat{x})$$

Anticipating that, the investor chooses the auditing intensity  $p$  to maximize the expected transfer to him net of auditing cost:

$$\max_p E\{\min_{\hat{x}} [pR(x, \hat{x}) + (1 - p)r(\hat{x})]\} - c(p).$$

Denote the maximizer of this expression  $p_s$ . The subscript  $s$  highlights the fact that the investor's choice of auditing intensity depends on the contract  $s$  accepted at period 0.

Back in period 0, the investor accepts the contract  $s$  if and only if the price he pays is no more than his discounted payoff at period 1. Due to our assumption that the contract is take-it-or-leave, the price  $q$  must in equilibrium be set to exactly break

the investor even:

$$q = \delta_i E\{\min_{\hat{x}}[p_s R(x, \hat{x}) + (1 - p_s)r(\hat{x})]\} - \delta_i c(p_s). \quad (3.1)$$

Now we calculate the entrepreneur's payoff then, which is the price  $q$  he receives plus his discounted payoff at period 1:

$$\begin{aligned} & q + \delta_e E[x] - \delta_e E\{\min_{\hat{x}}[p_s R(x, \hat{x}) + (1 - p_s)r(\hat{x})]\} \\ = & \delta_e E[x] + (\delta_i - \delta_e) E\{\min_{\hat{x}}[p_s R(x, \hat{x}) + (1 - p_s)r(\hat{x})]\} - \delta_i c(p_s) \end{aligned} \quad (3.2)$$

The equality is due to (3.1). The first term in (3.2) is the present value of the asset to the entrepreneur, which is exogenous. The second term is the expected gain from trade. And the last term is the present value of the investor's cost of auditing.

The following lemma concludes our first step of analysis, which establishes an upper bound for the entrepreneur's payoff at period 0:

**Lemma 3.3.1** *The entrepreneur's payoff at period 0 is at most  $\delta_e E[x] + \max_p\{(\delta_i - \delta_e)pE[x] - \delta_i c(p)\}$ .*

Proof:

$$\begin{aligned} & \delta_e E[x] + (\delta_i - \delta_e) E\{\min_{\hat{x}}[p_s R(x, \hat{x}) + (1 - p_s)r(\hat{x})]\} - \delta_i c(p_s) \\ \leq & \delta_e E[x] + (\delta_i - \delta_e) E\{[p_s R(x, 0) + (1 - p_s)r(0)]\} - \delta_i c(p_s) \\ \leq & \delta_e E[x] + (\delta_i - \delta_e)p_s E[x] - \delta_i c(p_s) \\ \leq & \delta_e E[x] + \max_p\{(\delta_i - \delta_e)pE[x] - \delta_i c(p)\} \end{aligned}$$

The second inequality is due to the feasibility requirements  $R(x, 0) \leq x$  and  $r(0) = 0$ . This concludes the proof.

Let  $V = \max_p \{(\delta_i - \delta_e)pE[x] - \delta_i c(p)\}$ , and  $p^*$  to be the corresponding maximizer. Note that  $V$  and  $p^*$  only depend on the primitives, not on any particular contract. And since the first term is linear in  $p$ , and  $c'' > 0$ , the maximizer  $p^*$  is unique.

This result is intuitive. At period 1, the entrepreneur will try her best to minimize the actual transfer to the investor through her reporting strategy. One of her feasible strategies is to always claim that she has nothing to transfer:  $\hat{x} \equiv 0$ . If so, for each possible realization of cash flow  $x$ , the investor gets nothing if he does not observe the true  $x$ , and if he does, he can at most take over the whole  $x$ . Thus, the expected maximum amount of transfer that can be enforced in period 1 is  $px$  if auditing intensity  $p$  is chosen. In addition, the entrepreneur may have other strategies that yield even less actual transfer. Therefore, at period 0, the expected gain from trade net of discounted auditing cost is at most  $\max_p \{(\delta_i - \delta_e)pE[x] - \delta_i c(p)\} = V$ .

Next, in our second step of analysis, we prove our main result that an equity contract achieves the upper bound for the entrepreneur's payoff established in Lemma 3.3.1.

**Proposition 3.3.1** *The contract  $s^* \triangleq (r(\hat{x}) = p^* \hat{x}, R(x, \hat{x}) = \begin{cases} p^* x, & \text{if } x = \hat{x} \\ x, & \text{if } x \neq \hat{x} \end{cases}, q = \delta_i \{p_{s^*} E[x] - c(p_{s^*})\})$ , with  $c'(p_{s^*}) = E[x]$ , maximizes the entrepreneur's payoff.*

Proof: At period 1, given this contract and any positive auditing intensity  $p$ , for each realization of  $x$ , the entrepreneur would choose  $\hat{x} = x$  if  $p \geq p^*$  and  $\hat{x} = 0$  otherwise.

The investor's optimal choice of auditing intensity,  $p_{s^*}$ , cannot be greater than  $p^*$ . Otherwise, we have

$$\begin{aligned}
& E\{\min_{\hat{x}}[p_{s^*}R(x, \hat{x}) + (1 - p_{s^*})r(\hat{x})]\} - c(p_{s^*}) \\
&= E[p_{s^*}x + (1 - p_{s^*})x] - c(p_{s^*}) \\
&= p^*E[x] - c(p_{s^*}) < p^*E[x] - c(p^*) \\
&= E\{\min_{\hat{x}}[p^*R(x, \hat{x}) + (1 - p^*)r(\hat{x})]\} - c(p^*)
\end{aligned}$$

This contradicts the optimality of  $p_{s^*}$ .

Since  $p_{s^*} \leq p^*$ , the entrepreneur always claims  $\hat{x} = 0$ , and the investor chooses  $p_{s^*}$  such that  $c'(p_{s^*}) = E[x]$ . The entrepreneur's period 0 payoff is then:

$$\begin{aligned}
& \delta_e E[x] + (\delta_i - \delta_e) E\{\min_{\hat{x}}[p_{s^*}R(x, \hat{x}) + (1 - p_{s^*})r(\hat{x})]\} - \delta_i c(p_{s^*}) \\
&= \delta_e E[x] + (\delta_i - \delta_e) E[p_{s^*}x + (1 - p_{s^*})0] - \delta_i c(p_{s^*}) \\
&\geq \delta_e E[x] + \delta_i E[p^*x + (1 - p^*)0] - \delta_i c(p^*) - \delta_e E[p_{s^*}x] \\
&\geq \delta_e E[x] + \delta_i E[p^*x] - \delta_i c(p^*) - \delta_e E[p^*x] = V
\end{aligned}$$

It reaches the upper bound established in Lemma 3.3.1. The first inequality results from the optimality of  $p_{s^*}$ , and the second inequality is due to  $p_{s^*} \leq p^*$  and  $c' > 0$ . This concludes the proof.

In the contract proposed in Proposition 3.3.1, for each realization of cash flow  $x$ , the *de jure* transfer is  $p^*x$ , and the *de facto* transfer is  $p_{s^*}x$ . Both of them take the form of an equity. This contrasts the result of Gale and Hellwig (1985), in which the

optimal contract is debt. Why is this the case? Note that there is a time-consistency issue with the entrepreneur. At period 1, she tries her best to minimize her transfer to the investor. But at period 0, she wants to maximize the transfer in order to enjoy as much gain from trade as possible. When designing the contract at period 0, she understands that the punishment for lying when caught is her only commitment device: Whatever auditing intensity the investor will choose, letting him take over all the cash flow if she is caught lying maximizes the transfer enforceable at period 1 and credible at period 0. Here, the non-contingency of the auditing technology plays the key role: since the auditing intensity  $p$  is constant over all contingencies, the enforceable transfer for each realization of cash flow  $x$  is  $px$ , which takes the form of an equity. When designing the contract, it is irrational to make the de jure transfer less than the enforceable transfer  $px$ , because that incentivizes the entrepreneur to tell the truth at period 1 and reduces the transfer. And the easiest way to make the de jure transfer greater than the enforceable transfer for each realization of cash flow  $x$  is to make it also an equity, with the proportion to the investor greater than the expected auditing intensity.

Note that this conclusion is robust, in the sense that it does not rely on the distribution of cash flow  $x$  or on the functional form of the auditing cost function  $c(\cdot)$ .



## 3.4 Conclusion

This chapter studies the contract design implication of routine auditing with a model resembling Gale and Hellwig (1985) otherwise. We prove that an equity contract is optimal. This contrasts the result of existing costly state verification literature and highlights the role played by the state-contingency of auditing decisions.

## 3.5 References

1. Axelson U. 2007. "Security Design with Investor Private Information", *The Journal of Finance*, 62 (6), 2587-2632.
2. Carroll G. 2015. "Robustness and Linear Contracts", *American Economic Review*, 105 (2), 536-563.
3. Chassang S. 2014. "Calibrated Incentive Contracts", *Econometrica*, 81 (5), 1935-1971.
4. Diamond P. 1998. "Managerial Incentives: On the Near Linearity of Optimal Compensation", *Journal of Political Economy*, 106 (5), 931-957.
5. Gale D. and M. Hellwig. 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem", *The Review of Economic Studies*, 52 (4), 647-663.
6. Holmstrom B. and P. Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica*, 55 (2), 303-328.

7. Mookherjee D. and I. Png. 1989. "Optimal Auditing, Insurance, and Redistribution", *Quarterly Journal of Economics*, 104, 399-415.
8. Krasa A. and A. Villamil. 2000. "Optimal Contracts When Enforcement is a Decision Variable", *Econometrica*, 68, 119-134.
9. Townsend R. 1979. "Optimal contracts and Competitive Markets with costly State Verification", *Journal of Economic Theory*, 21, 265-293.
10. Winton A. 1995. "Costly State Verification and Multiple Investors: The Role of Seniority", *Review of Financial Studies*, 8, 91-123.