Dynamic Compensation and Investment with Limited Commitment

by

Felix Zhiyu Feng

Department of Economics
Duke University

Date: ___________________________

Approved:

____
Curtis R. Taylor, Supervisor

____
Adriano Rampini

____
Attila Ambrus

____
S. Viswanathan

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Dynamic Compensation and Investment with Limited Commitment

by

Felix Zhiyu Feng

Department of Economics
Duke University

Date: ____________________________
Approved:

____
Curtis R. Taylor, Supervisor

____
Adriano Rampini

____
Attila Ambrus

____
S. Viswanathan

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2014

Abstract

In this dissertation I study the role of limited commitment in dynamic models. In the first part, I consider firms that face uncertainty shocks in a principal-agent setting but have only limited ability to commit to long-term contracts. Limited commitment firms expedite payments to their managers when uncertainty is high, a finding that helps to explain the puzzling large bonuses observed during the recent financial crisis. In the second part, I examine a dynamic investment model where firms invest in a risky asset but cannot hedge the risk of their investment because they lack the ability to commit to future repayments of debt. Once firms have access to exogenous supplies of risk free assets, they may on an aggregate level invest more in the risky asset, because risk free technology allows them to increase in wealth in equilibrium. This result helps to explain the asset price booms in emerging countries when they experience substantial capital outflow.

To Caroline

Contents

List of Figures

Acknowledgements

I am deeply indebted to Attila Ambrus, Barney Hartman-Glaser, Adriano Rampini, Curtis Taylor, and S. Viswanathan for invaluable guidance. I owe my thanks to Hengjie Ai, Philip Bond, Michael Brolley, Peter Cziraki, Brendan Daley, Brett Green, Dirk Jenter, Andrey Malenko, Ernst Maug, Thomas Noe, Vincenzo Quadrini, Philipp Sadowski, Andrei Shleifer, Ajay Subramanian, Ed Tower, Felipe Varas, Dimitri Vayanos, Paolo Volpin, Amy Ward, Daniel Yi Xu, Ming Yang for the discussions during various stages of writing this dissertation. I also thank other seminar and conference participants at Duke University, Georgia State University, University of Notre Dame, University of Oxford, University of Washington, the 2013 Trans-Atlantic Doctoral Conference and the 2013 EFA meetings for helpful comments. All errors are my own.

1

Introduction

Contracts are important in achieving efficient resource allocations. The effectiveness of contracts depends on transacting parties' ability to commit to the contractual terms. Research usually assumes that contracts are by and large enforceable due to the existence of a legal system. This assumption however may not always hold, as limited ability of contractual commitment actually prevails in many transacting relationships and in many markets. In this dissertation, I consider two cases in corporate finance where contracts may not be perfectly enforced: one, a compensation contract relationship, where firms may not be able to commit to contract termination time because of at-will employment and firms' freedom to liquidate. Two, firms in emerging economies may not be able to commit to future debt repayments, because of imperfect legal systems.

In the first chapter I study optimal dynamic compensation with limited commitment to contract termination, when firms are subject to uncertainty shocks. I analyze a continuous-time dynamic principal-agent model with private effort and regime switching in cash flow volatility and characterize the optimal managerial compensation and termination policy. In high volatility times, firms are forced to expedite payments to managers because sufficient deferred compensation is no longer credible. At the same time, contract length shortens, that is, termination becomes more likely. This relation between the timing of payments and expected contract length may explain the sizeable cash bonuses observed in crises times. In contrast, with full commitment firms defer compensation more when volatility is high.

In the second part I examine limited commitment in a dynamic investment model in the context of global financial integration. I shows that in emerging economies where contractual commitment is imperfect, capital outflows from emerging economies after financial integration can lead to simultaneous increases in the savings rate and in domestic asset prices. Under autarky, firms in emerging economies invest in risky capital while facing a borrowing constraint that creates a need for precautionary savings. Financial integration provides firms with access to foreign risk-free assets and results in two effects: a substitution effect, whereby firms divert some investments to foreign assets and cause capital outflows; and a wealth effect, whereby they grow richer in equilibrium and thus demand more domestic capital. Savings gluts and asset price booms occur when the wealth effect dominates. The increases in savings and asset prices are inefficiently high relative to the socially optimal level and can be amplified by heterogeneity in productivity among domestic firms.

2

Dynamic Compensation and Uncertainty Shocks

Managerial compensation is among the most controversial issues brought out by the recent financial crisis. How compensation is designed matters for providing managers with proper incentives, especially when the outcome of managerial effort bears great uncertainty. Because individuals are forward-looking, their valuation of compensation depends on the future payments they can expect and therefore hinges on firms' ability to commit to making these payments. The theoretical literature on financial contracts usually assumes that firms have perfect ability to commit, which is not always the case in practice.

A simple example illustrates why assuming full commitment power is potentially restrictive. In a canonical principal-agent model, pay-for-performance is used to solve agency problems. The principal rewards the agent through payments such as bonuses when performance measures are sufficiently strong and punishes him usually through contract termination when performance measures are weak. In other words, firms' commitment is two-fold: commitment to make payments when due and commitment to retain managers until a termination condition is triggered. The latter type of commitment is generally infeasible in practice given the prevalence of at-will employment. Under US labor law, firms can fire employees without having to establish just cause or give warning. Firms can also liquidate anytime, after which they are no longer liable for any future compensation promised to employees.

In this chapter, I study the optimal compensation contract where firms face uncertainty shocks but cannot credibly guarantee future payments to managers because they can walk away from the contract at any time. I refer to this ability to unilaterally terminate contracts as limited commitment and explore its implications for the dynamics of compensation, in particular crisis-time incentive payments. In the recent financial crisis, managerial compensation drew wide attention from the general public and academic researchers alike when it was revealed that huge losses of company wealth notwithstanding, many bankers and executives still received substantial compensation. Most notable are the bonuses paid out by financial firms, for example Merrill Lynch, which paid out a total of \$3.6 billion in bonuses in the 2008 fiscal year despite having suffered losses of \$27 billion; and Citigroup, which paid out \$5.3 billion in bonuses after a \$27.7 billion loss[1].

The controversially large compensation during the recent crisis warrants further investigation of firms' compensation practices in the face of uncertainty shocks. Uncertainty lies at the heart of financial crises, as pointed out by the growing research showing that changes in real business outcomes are driven by changes in underlying investment risks[2]. Financial firms, among other businesses, are the most sensitive to uncertainty shocks and therefore rely heavily on incentive pay, which makes up the majority of overall firm cost as well as total employee compensation. Incentive pay, however, is not determined by uncertainty alone. It also depends on firms' ability to

[1] Wall Street Journal: Wall Street Compensation–'No Clear Rhyme or Reason'. July 30th, 2009. More detailed statistics can be found in the press release of the New York State Comptroller (2009) as well as in Frydman and Jenter (2010) and Kaplan (2012)

[2] See for instance Bloom (2009), Brunnermeier and Sannikov (2012), Atkeson, Eisfeldt, and Weill (2013), Di Tella (2013), He and Krishnamurthy (2013a)

commit to making the promised incentive payments. Moreover, limited commitment is more of an issue during crisis times when uncertainty is high, since firm liquidation and managerial turnover are more frequent and concern over firms' ability to abide by their commitments is greater.

I argue that large incentive pay in high uncertainty crisis times can result from optimal compensation contracts, under limited commitment. The engine of analysis is a continuous time principal-agent model in which the agent, representing a manager, can choose an effort level that is unobservable to the principal. Given this potential for moral hazard, the principal's optimal contract requires that the firm incentivize the manager by promising future payments that depend on performance. These payments are made in the form of cash bonuses after managers exceed a certain performance benchmark. However, when firm value is low and promised future payment to managers high, firms with only limited power of commitment will be tempted to terminate managers' contracts, thereby avoiding all future obligations. Faced with the possibility of losing future compensation, managers will not agree to a prolonged delay of payments. This relationship between payment timing and contract length leads firms to compensate their managers more immediately in order to maintain proper incentives.

I model uncertainty shocks through stochastic regime switching between low and high volatility states, representing normal and crisis times. Under regime switching, firms optimally allocate managerial deferred compensation until the marginal value before the uncertainty shock is equal to the marginal value after the shock. These important dynamics are absent from simple comparative statics, which implicitly hold managerial deferred compensation constant when comparing different volatility levels. However, the capacity to defer compensation is constrained by how much future payment firms can credibly pledge. Without the ability to commit to the timing of contract termination, firms are forced to substitute future payments with immediate payments whenever firm value is low and termination more likely, resulting in more cash bonuses being paid once the crisis state obtains. These payments should not be confused with a "reward." In fact, the lifetime present value managers derive from a contract that results in more immediate payments during crises is lower. This is because managers are simultaneously subject to a higher probability of termination. In contrast, firms with full commitment defer compensation more when volatility is high, and managers may be consequently rewarded with a higher present value of total compensation. These predictions are generally in line with empirical evidence such as Peters and Wagner (2013), who show that increases of market volatility lead to more forced managerial turnover, which in turn result in higher managerial compensation.

Despite many studies on managerial compensation, direct analysis of why compensation (in particular the bonuses paid by financial firms) during crisis time remains high is rare. Several theories attempt to rationalize the escalating use of incentive compensation and high-powered contracts for financial firms with firms' competition for managerial talent. Glode, Green, and Lowery (2012) build a model where firms over-invest in financial experts to gain strategic advantage in subsequent trading. Axelson and Bond (2012) introduce large compensation for employees of financial firms through the size of capital they manage and argue why competition among employees does not depress compensation. Bénabou and Tirole (2013) model how competition for productive managers interacts with firms' incentive structure through a multitasking and screening model and explore the consequences of regulatory policies such as bonus caps, a feature also discussed in Bijlsma, Boone, and Zwart (2012), Thanassoulis (2012), Bond and Glode (2013), Glode and Lowery (2013). While explaining the unusually and increasingly high compensation earned by the employees of the financial sector, these existing theories do not answer whether and why such compensation remains high when market fundamentals deteriorate. Their argument also relies on the heterogeneity and scarcity of talent in the financial industry, which is more applicable to senior executives and CEOs. However, empirical evidence from Oyer (2008), Kaplan and Rauh (2010), Philippon and Reshef (2012). show that high level incentive payments extend beyond a handful of top managerial elites, who are arguably more skillful and can earn a larger premium. In contrast, agents in my model are homogenous and large payments in the form of cash bonuses are still possible even if the market is highly volatile because firms choose the level of immediate compensation optimally depending on their credibility to promise future payments.

This chapter makes several contributions to the literature of contract theory, managerial compensation, and corporate governance. On the modeling side, to my knowledge, this is the first paper that jointly considers agency, limited commitment, and regime switching. It is also one of the first to examine the relationship between compensation and volatility, whereas most of the extant research on managerial compensation has so far focused on profitability[3]. Meanwhile, this chapters theoretical results generate testable empirical hypotheses: conditional on negative uncertainty shocks, commitment-constrained firms make larger immediate payments and have higher managerial turnover relative to unconstrained firms. These hypotheses have implications for the understanding and evaluation of firms' governance. While the empirical literature on corporate governance generally takes low total compensation and high pay-for-performance sensitivity as indicative of good governance, this chapter shows the importance of considering the level and structure of compensation under the context of market uncertainty. Total compensation and pay-for-performance sensitivity are sensible proxies for firm governance only when firms have no commit-

[3] Except for studies on the sensitivity of pay-for-performance, such Lambert and Larcker (1987), Aggarwal and Samwick (1999), and Core and Guay (2002). See Prendergast (2002) for a survey on this topic. These studies usually do not focus on the correlation between volatility and the the level of pay itself

ment issue, which may not always be true, especially during spells of high volatility.

Results of this chapter put forth a caveat to the popular perception that the high compensation observed in the recent financial crisis is a sign that managers are entrenched and the current compensation structure largely suboptimal (or even corrupt). This public notion has motivated political activism to regulate and reform compensation practices[4]. However, without taking into account firms' commitment ability, policies intended to align managers' incentives with those of investors could actually backfire. For instance, TARP recommends deferring compensation to executives and the Dodd-Frank Act gives shareholders the right to disapprove any golden-parachute compensation to executives. However, executives' valuation of deferred compensation depends crucially on their assessment of firms' ability to commit to making future payments. With escalating uncertainty it is difficult for firms to maintain managers' confidence, so they must be paid with more immediacy. Thus, during times of crisis, following the recommendation to defer compensation and restrict retention payments would actually do more harm than good by undermining managerial incentives.

Using a novel approach, I also provide a justification for the limited commitment constraint based on the security implementation of optimal contracts. The implementation involves standard securities such as equity and debt and hence potential tension between their holders. Following an uncertainty increase, the face value of long-term debt must decline, implying the redemption of debt which entails a wealth transfer from equity holders. Under high volatility, firm value is low, in which case equity holders may find it optimal to default rather than recall the debt. Therefore, in the implementation of the contract, assuming that the principal has full ability to commit to the agent is equivalent to assuming that equity holders have full abil-

[4] See, for instance, the Troubled Asset Relief Program (TARP): Executive Compensation Rules & Guidance and the Dodd-Frank Wall Street Reform and Consumer Protection Act

ity to commit to maintaining a certain capital structure. The latter assumption is largely unrealistic as equity holders usually can default on debt without considering the effect on the firm as a whole.

Finally, this model also reveals the possibility that different levels of effort from the agent could be implemented by the optimal contract under different levels of cash flow uncertainty. As uncertainty increases, the value of an incentive compatible contract decreases. When the uncertainty shock is bad enough, the optimal contract can shift from being incentive compatible to one that allows shirking. The agent stops receiving cash payments and is instead compensated through his private benefit from shirking, and due to lower managerial effort, expected cash flow falls. Thus, if a crisis is substantially severe, managers will stop receiving bonuses. This, however, should be of little comfort because it implies managerial indifference and correspondingly low productivity. This finding also suggests the potential edogeneity between profitability and volatility, which challenges empirical studies on executive compensation that treat the two as independent.

The remainder of the chapter is organized as follows: Section 2.1 briefly reviews the related literature. Section 2.2 describes the main model with fixed uncertainty level and solves for the optimal contract under limited commitment. Section 2.3 introduces uncertainty shocks and derives the model's implications on dynamic compensation. Section 2.4 discusses two major extensions: security implementation and shirking in equilibrium. Section 2.5 concludes.

2.1 Related Literature

This chapter builds on the rich line of literature regarding optimal compensation in dynamic environments. Early work on dynamic agency problems such as Holmstrom and Milgrom (1987), Spear and Srivastava (1987) points out that moral hazard calls for the principal to delay payments to the agent until some measure of performance is observed. The agent's continuation value, which is the net present value of all future income he expects to receive from the contract, is a powerful tool that can be used to characterize the optimal recursive contract. In more recent years the development of stochastic dynamic control has lead to the rise of continuous time agency models which, compared to their discrete time counterparts, yield highly tractable solutions. Sannikov (2008) derives the seminal martingale method for continuous time dynamic agency models. Since then the method has been applied to models with different cash flow processes, such as arithmetic Brownian motion in DeMarzo and Sannikov (2006), geometric Brownian motion in He (2009), and point processes in Biais et al. (2010).

In this chapter, the cash flow process and the moral hazard problem are adopted from DeMarzo and Sannikov (2006), where a manager with private effort contracts with a principal who has a project yielding continuous cash flows with increments following an arithmetic Brownian motion. The manager in DeMarzo and Sannikov (2006) is protected by limited liability, while the principal has full commitment power. The incentive compatible contract requires that the manager is not paid until his present value from the contract exceeds a payment boundary chosen optimally by the principal according to the state of the economy. During crisis times when cash flow uncertainty soars, the optimal contract with full commitment would postpone the payment boundary even farther and the manager would be less likely to receive cash payments in any given period. Such a prediction does not square with the sizable bonuses paid by many distressed financial firms in the recent crisis. This observation motivates relaxation of the strong assumption of full commitment power by the principal.

Indeed, one of the two key assumptions at the heart of this chapter is the principal's lack of commitment to future payments. This assumption alone has been explored in prior work: Thomas and Worrall (1988, 1994), Abreu et al. (1990) and

Ray (2002) discuss the general methodology that solves the optimal contract through discrete repeated games and find that the equilibrium can be characterized by two state variables: the respective continuation values of the two contracting parties. The key is to find the set of values of the state variables for which neither party has the incentive to default in equilibrium or, commonly in the literature, for which the contract is self-enforcing. Recent studies of Grochulski and Zhang (2011), Miao and Zhang (2013), and Ai and Li (2013) develop the continuous time version of such self-enforcing contracts, where a risk-averse manager runs a project with random cash flows and must hedge his exposure to the cash flow risk with a principal who can be risk averse or risk neutral. These studies provide some technical methodology from which this chapter is adapted. Compared to the discrete time repeated games, which generally involve multiple equilibria, continuous time cash flow models usually produce an unique optimal contract[5].

This chapter differs from the aforementioned studies involving limited commitment in a number of dimensions. Most importantly, The majority of the limited commitment literature focuses on the optimal risk sharing contract between a risk-neutral principal and a risk-averse agent. In such environments , termination is never part of an optimal compensation scheme. By contrast, I study a setting with a risk-neutral agent who is protected by limited liability. In this environment, incentives are optimally provided in part through threat of contract termination. This is a critical feature for my study given the high frequency of managerial turnover in times of crisis. Assuming risk neutrality and constant outside options also greatly lowers the technical hurdles for solving the optimal contract.

In addition to the analysis of compensation contracts, both agency and limited

[5] One implicit difference between the two frameworks is the timing of the players' actions. Under the former framework, players usually move simultaneously. In the latter models, the principal makes payment decisions after observing the cash flow signal, which avoids the coordination problems that arise from a repeated game with simultaneous actions.

commitment have been widely applied to a broad range of other topics in finance. For instance, agency based dynamic models serve as a convenient tool for analyzing firm financing, investment, and capital structure, most notably by Albuquerque and Hopenhayn (2004), Clementi and Hopenhayn (2006), DeMarzo and Fishman (2007), Rampini and Viswanathan (2013a). Likewise, limited commitment has been applied extensively to areas in, for example, macro and international economics[6]. In Section 5 I explore explicitly the capital structure implied by implementing the optimal contract with common securities. I show that the implementation assumes implicitly that shareholders can commit to a certain capital structure, which is widely known to be implausible. This equivalence serves as evidence that contracts with limited commitment are potentially more prevalent than previously thought.

Several studies attempt to integrate agency problems into the limited commitment framework, such as Atkeson (1991) and Levin (2003) in discrete time and more recently Fong and Li (2012), Grochulski and Zhang (2013), and Opp and Zhu (2013) in continuous time. The structure in Levin (2003) has become wildly known as the relational contract. My paper differs from relational contracts in that I allow a one-time contract termination, which can be naturally considered as firm liquidation, as the only way in which the principal cannot credibly commit. Conditional on a firm remaining in operation, the principal can commit to all the payments specified in the contract. This allows long-term contracts as a viable means of overcoming agency problems despite the imperfect enforcement condition. Furthermore, most of the existing models are set in a stationary environment. As described in the introduction, a stationary environment is insufficient for characterizing the dynamics of compensation when the state of the economy shifts, as firms will adjust their compensation in a state-contingent manner to optimally allocate the value of compensation. In

[6] To name a few, Bulow and Rogoff (1989), Kehoe and Levine (1993), Alvarez and Jermann (2000), Kehoe and Perri (2002), Cooley et al. (2004)

contrast to the existing literature, my model involves regime switching and therefore provides stronger implications for the dynamics of compensation during crisis versus normal times.

This regime switching technique for continuous time models is adopted from Hoffmann and Pfeil (2010), Piskorski and Tchistyi (2010), Bolton, Chen, and Wang (2013), DeMarzo, Fishman, He, and Wang (2012). My paper is closely related to Hoffmann and Pfeil (2010), which also adopts the DeMarzo and Sannikov (2006) framework with private effort and an arithmetic Brownian motion cash flow process. This paper differs from the others which focus mainly on stochastic regimes of profitability by considering regimes of cash flow volatility. Although the key results are generally replicable with stochastic profitability, volatility is more suitable for analyzing the compensation of financial firms which predominantly consists of contingent incentive payments rather than fixed salary payments.

To my best knowledge, this paper is the first to combine moral hazard and limited commitment with regime switching in a continuous time framework. This innovation allows me to rationalize several important crisis time observations regarding managerial compensation that have largely escaped understanding before now.

2.2 Model

In this section I describe the model in a standard principal-agent environment with only one regime of uncertainty. I first solve the optimal contract assuming that the principal has full commitment power. After discussing the limitation of this strong assumption, I then solve the optimal contract but suppose limited commitment by the principal.

2.2.1 Basic Environment

Time is continuous. A principal, representing a firm, must hire an agent, representing a manager, to run a project. Both the principal and the agent are risk neutral. The cash flow Y_t of the project follows

$$dY_t = \mu(e_t)dt + \sigma dZ_t \ ,$$

where Z_t is a standard Brownian motion; $\mu(e_t)$ is the expected growth rate of cash flow depending on the agent's effort. σ is the cash flow volatility, which measures the level of uncertainty a firm faces when taking on such a project.

The agent controls the cash flow growth rate by choosing a binary effort level $e_t \in \{\overline{e}, \underline{e}\}$, representing "working" and "shirking", respectively. I assume $\mu(\overline{e}) = \mu$ and $\mu(\underline{e}) = \mu - C$ where $C > 0$, that is, shirking results in lower expected cash flow. However, the agent enjoys a private benefit λC whenever he shirks, where $\lambda \in (0, 1]$ measures the degree of agency problem in this model. Effort is private to the agent: the principal can observe Y_t but not e_t. The principal discounts future cash flows by r and the agent by $\gamma > r$, so the agent is more impatient.[7]

For now, I assume that the principal can commit to any contract once it is signed, but the agent is protected by limited liability with an outside option whose value R is normalized to 0. Limited liability implies that payments to the agent must be non-negative and that contract termination is necessary for incentive provision. The principal has an outside option L which she receives whenever the contract is terminated. Both the principal and the agent take no further action after the contract termination, which eliminates reputation concerns. One can interpret contract

[7] The asymmetry of discount rates is essential for a non-trivial incentive compatible contract to exist in this type of model. Because of the agent's private effort, the principal must delay cash payments to the agent until his cumulative performance exceeds a certain threshold. Having the agent discount future cash income heavier ensures that providing incentive through such delaying is costly for the principal so the principal will not want to delay payments forever. However, once I impose the principal's commitment constraint, this additional constraint leads to the existence of an optimal contract even for the case where r=γ, which I describe in Section 4

termination here as a firm's liquidation and exit from the market permanently or, equivalently, as a firm's replacement of managers, where L is the normalized net profit from contracting with a new manager.

I also assume for most part of this chapter that the principal prefers to induce the agent to work. This is true as long as C, the cost of shirking, is high enough. In Section 5 I discuss how sufficiently large uncertainty shocks can allow shirking in equilibrium because incentive provision is too costly when uncertainty is high.

Let $\mathscr{F}_t$ be the filtration generated by the cash flow history. A contract which specifies a compensation process $\{I_t\}_{t\geqslant 0}$ from the principal to the agent, a termination time τ, and a recommended effort process e_t defines the agent's continuation utility W_t:

$$W_t = E\left[\int_t^\tau e^{-\gamma(s-t)}\left(dI_s + \lambda C \mathbb{1}_{\{e_t=\underline{e}\}}dt\right)|\mathscr{F}_t\right] .$$

where $\mathbb{1}_{\{e_t=\underline{e}\}}$ is an indicator function that takes value 1 if $e_t = \underline{e}$ and zero other wise. W_t simply measures the present value of all expected future payments and can be intuitively interpreted as the agent's "wealth". Similarly, the contract defines the principal's valuation of the project V_t which is the expectation of total future cash flow minus the payment to the agent plus the liquidation value when the contract is terminated.

$$V_t = E\left[\int_t^\tau e^{-r(s-t)}\left(dY_t - dI_s\right) + e^{-r(\tau-t)}L|\mathscr{F}_t\right] ,$$

So far the basic environment is identical to the one in DeMarzo and Sannikov (2006). I therefore briefly review their solution for the optimal contract to offer a benchmark for later comparison. Using the martingale method developed by Sannikov (2008), there exists $\mathscr{F}_t$ measurable processes β_t such that W_t evolves according to

$$dW_t + dI_t = \gamma W_t - \lambda C \mathbb{1}_{\{e_t=\underline{e}\}}dt + \beta_t(dY_t - (\mu - C\mathbb{1}_{\{e_t=\underline{e}\}})dt) . \qquad (2.1)$$

Equation (2.1) means the principal can compensate the agent with either immediate payments dI or future payments dW. Payments are sensitive to the agent's performance, i.e. the realized cash flows dY_t, and the sensitivity is measured by β_t. DeMarzo and Sannikov (2006) shows that the contract is incentive compatible if and only if $\beta_t \geqslant \lambda$ for all $t < \tau$. Intuitively, β_t represents the agent's "skin in the game" in the incentive compatible contract or the sensitivity of the agent's continuation utility to the realized cash flow. The sensitivity must be no less than λ which is the amount of private benefit the agent can derive per unit of cash flow he looses from shirking.

Under the optimal incentive compatible contract the principal's valuation of the project V_t is a function of the agent's continuation utility W_t. The function is denoted as $V(W_t)$ and is referred to as firm value, since the principal represents a firm in this contracting relationship. The principal must earn an instantaneous return r under the optimal contract through the expected cash flow and the expected change rate of her valuation of the project, that is

$$V(W_t) = E\left[dY_t + dV(W_t)\right] \ . \tag{2.2}$$

Applying Ito's lemma to (2.2) and imposing $e_t = \overline{e}$ on equation (2.1), $V(W_t)$ is characterized by the following Hamiltonian-Jacobian-Bellman (HJB) equation:

$$rV(W_t) = \max_{\beta_t \geqslant \lambda} \mu + \gamma W_t V'(W_t) + \frac{1}{2}\beta_t^2 \sigma^2 V''(W_t) \ . \tag{2.3}$$

The principal maximizes her HJB equation by choosing β_t optimally. The concavity of the value function implies $\beta_t = \lambda$ for all t as long as the principal induces working as the equilibrium effort. On the one hand, $\beta_t \geqslant \lambda$ is necessary for incentive reasons. On the other hand, it is not optimal to set β above λ since higher sensitivity increases the likelihood of contract termination which is also a costly action for the principal.

The principal's value function must also satisfy three additional conditions which define two boundaries for W_t: the termination boundary and the payment boundary. First, the contract is terminated when W_t hits R. The principal cannot provide further incentives by lowering the agent's continuation utility any more since the agent is protected by limited liability. The principal receives the liquidation value L at the time of contract termination, so $V_s(R) = L$. Secondly, since the principal can always make a lump sum transfer of dI to the agent, the principal's valuation of the project satisfies $V(W) \geqslant V(W - dI) - dI$. Equivalently this means $V'(W) \geqslant -1$, that the shadow value of the agent's continuation utility to the principal should not be lower than the cost of an instant cash payment. Defining $\overline{W} = \inf\{W | V'(W) = -1\}$ as the *payment boundary*. $\overline{W}$ satisfies the "smooth pasting" condition $V'(\overline{W}) = -1$. The agent will receive instant cash payment of size $W_t - \overline{W}$ once $W_t > \overline{W}$, which brings W_t back to $\overline{W}$ immediately. Finally, the payment boundary $\overline{W}$ is optimally chosen by the principal, which implies the "super contact" condition: $V_s''(\overline{W}) = 0$.

The following lemma, which is also Proposition 1 in DeMarzo and Sannikov (2006), summarizes the optimal incentive compatible contract in this setting:

Lemma 2.2.1. *The optimal incentive compatible contract, which maximizes the principal's payoff subject to delivering value W to the agent and satisfying the agent's limited liability, can without loss of generality be characterized by a value function $V(W)$ and a payment boundary $\overline{W}$ such that*

$$rV(W) = \mu + \gamma W V'(W) + \frac{1}{2}\lambda^2\sigma^2 V''(W) \,,$$

with boundary conditions $V(R) = L$; $V'(\overline{W}) = -1$; $V''(\overline{W}) = 0$. Immediate cash payment of size $W_t - \overline{W}$ is made for all $W_t > \overline{W}$. When $W_t = R$, the contract is terminated.

Under the optimal contract, the agent receives compensation $dI = \max(W_t -$

$\overline{W}, 0)$, which is always non-negative because of his limited liability and strictly positive whenever his continuation utility W_t exceeds the payment boundary $\overline{W}$. This compensation scheme resembles cash bonuses that managers in practice receive for good performance, where $\overline{W}$ represents the "bonus hurdle" managers must clear before getting paid. Hence, this model is suitable for studying the compensation of financial firms, in which cash bonuses make up the bulk of total employee compensation. Note that Lemma 2.2.1 concludes that any optimal incentive compatible contract can without loss of generality be characterized by a contract involving incentive pay that resembles cash bonuses, a natural means of incentive provision in this setting. Thus, in the remainder of this chapter, I refer to the cash payment dI as bonuses. Keep in mind that it is representative of incentive pay in general, which in the current model can be contracted on and hence to which the principal can commit.

The optimal contract is associated with welfare losses due to moral hazard. The efficient allocation calls for the principal and the agent to split the maximal surplus generated by running the project permanently; that is, the agent's compensation and the principal's payoff satisfy $V(W) + W = \mu/r$, which marks the Pareto frontier of this model in the absence of moral hazard. However, when moral hazard is present, the principal must design incentive compatible contracts featuring delayed payments to the agent. Substituting boundary conditions $V'(\overline{W}) = -1$ and $V''(\overline{W}) = 0$ into the principal's HJB equation yields $rV(\overline{W}) + \gamma\overline{W} = \mu$, which is the "second best" frontier below the Pareto efficient frontier. This is because the agent is more impatient and some surplus is lost in the optimal incentive compatible contract with delayed payment. The "second best" frontier becomes a critical boundary when considering the optimal contract with principal's limited commitment, which is introduced in the following subsection.

2.2.2 The Commitment Constraint

So far, the structure of the optimal contract characterized in Lemma 2.2.1 relies on the principal's commitment to all future payments once the contract is signed. However, before the agent's continuation utility W hits the payment boundary $\overline{W}$, the agent is not actually paid. His continuation utility measures the present value of the total amount of payment he expects to receive in the future, only if the principal honors the contract. Just as the agent is tempted to quit his job when W approaches his reservation utility R, the principal will likewise be tempted to exercise her outside option, which in this model is liquidating the project and receiving L, if the firm value V drops below the liquidation value before W reaches the payment boundary. If enforcement is not perfect and commitment becomes a binding constraint before the cash payment boundary is reached, the dynamics of the optimal contract will consequently be different.

To consider this impact, I assume that the principal can terminate the agent's contract anytime. As discussed earlier, this assumption of limited commitment on the part of the principal is more realistic, as firms generally are free to fire managers or liquidate projects at any time in practice. Once the contract is terminated, I assume both parties will receive the value of their outside options: L for the principal and R for the agent. This assumption sets this model apart from the other models in the relational contract literature in that termination time is the only aspect of the contract to which the principal cannot commit. Conditional on the continuation of the contract, the principal can still commit to all payments once the payment boundary is reached, suggesting the existence of long-term contracts although subject to a participation constraint from the principal[8].

[8] At this point the principal can also commit not to renegotiate the contract, even though renegotiation can be mutually beneficial. I discuss the renegotiation-proof contract under this setting in the Appendix

The assumption that both the principal and agent receive constant outside value whenever the contract is terminated is critical to obtaining a close-form solution. Existing models of limited commitment usually consider endogenous outside options, and common dynamic programming techniques usually do not apply directly. In my model the constant outside option assumption greatly simplifies the analysis by allowing the application of the Sannikov (2008) method to solve the moral hazard problem. Moreover, since the cash flow in this model follows an arithmetic Brownian motion, constant outside option value rules out strategic control of effort choices or firm size whenever the participation constraints are binding or close to binding.

I introduce here a heuristic approach that derives the optimal contract under principal's limited commitment by separating the commitment constraint from moral hazard, the other contractual friction in the model. First, suppose the agent's effort is observable to the principal and then the only contractual constraints are the principal's limited commitment and the agent's limited liability. Limited commitment implies a participation constraint for the principal:

$$V_t \geqslant L \ . \tag{2.4}$$

Combined with the agent's participation constraint $W_t \geqslant R$, they define a payoff space $\{(W,V)|W \geqslant R, V \geqslant L\}$ where, if the continuation value delivered by a contract falls into the space, the contract will not be terminated, i.e. the contract is *self-enforcing*.

Given the self-enforcing contracting space, consider now adding moral hazard. The martingale representation theorem implies the dynamics of W still follows equation (2.1), and the optimal contract still features a reflecting payment boundary $\overline{W}$. Combined with the principal and agent's participation constraints, these conditions define a space between $W \geqslant R$, $V \geqslant L$ and the "second best" frontier $rV + \gamma W = \mu$ in which the optimal contract value function must lie. The principal's valuation of the project V is a function $V(W)$ of the agent's continuation utility, where $V(W)$ satisfies the same HJB in equation (2.3).

The exact shape of the value function is pinned down by appropriate boundary conditions similar to those characterized in Lemma 2.2.1. Intuitively, at the termination boundary, firm value must match the liquidation value L. At the payment boundary, the marginal value of cash payment always equals the shadow value of the agent's continuation utility. This implies the "smooth-pasting" condition $V'(\overline{W}) = -1$ always holds. The last boundary condition depends on whether the value function crosses the "second best" frontier $rV(W) + \gamma W = \mu$ or the self-enforcing border $V(W) = L$ first. If the value function meets $rV(W) + \gamma W = \mu$ first, it immediately follows $V''(\overline{W}) = 0$ and the principal's commitment constraint is not binding. In contrast, if the value function reaches $V(W) = L$ first, then $V(\overline{W}) = L$ is the boundary condition that replaces $V''(\overline{W}) = 0$.

These boundary conditions are intuitive. By concavity of the value functions, if the payment boundary is such that both the limited commitment constraint and the "super contact" condition are slack, in other words $V(\overline{W}) > L$ and $V''(\overline{W}) < 0$, the principal can always achieve a higher value by postponing the payment further, until either condition becomes binding. If the commitment constraint is binding with a lower $\overline{W}$, the payment boundary is no longer optimally chosen, and the "super contact" condition is replaced with a physical boundary condition $V(\overline{W}) = L$. The reason why only firm value at the payment boundary turns out to matter under limited commitment is the combination of a concave value function, $V(R) = L$ on the left boundary, and $\overline{W}$ as a reflecting right boundary.

Let variables with a superscript L represent variables in the limited commitment environment, the following proposition summarizes the optimal contract. A formal verification theorem of the optimality of this contract is provided in the appendix.

Proposition 2.2.1. *The optimal contract under the principal's limited commitment constraint is characterized by a value function* $V^L(W)$ *and a payment boundary* $\overline{W}^L$*, such that*

$$rV^L(W) = \mu + \gamma W V^{L\prime}(W) + \frac{1}{2}\lambda^2\sigma^2 V^{L\prime\prime}(W) \ ,$$

subject to boundary conditions $V^L(R) = L$*;* $V^{L\prime}(\overline{W}^L) = -1$*; and*

$$V^{L\prime\prime}(\overline{W}^L) = 0, \ \textit{if} \ V^L(\overline{W}^L) \geqslant L \ ,$$

$$V^L(\overline{W}^L) = L, \ \textit{otherwise.}$$

As in the full commitment benchmark case, the principal's value functions $V^L(W)$ under limited commitment is also a concave function with an interior maximal point. The principal's commitment constraint becomes binding when W is high, in other words when the manager has accumulated adequate performance history. This is not counter-intuitive as W is a measure of the amount of future bonuses managers are owed, a form of debt of the firm induced by the labor contract. Higher debt in the form of bonuses lowers the share of profit investors can earn from investing in the firm. When it becomes large enough, investors will find it optimal to default like they would with any other form of debt. In that sense, the commitment constraint can also be motivated as an upper bound for the operating leverage of the firm.

Several immediate implications can be made by comparing the limited commitment contract and the full commitment contract. The following conclusions can be shown straightforwardly:

Corollary 2.2.1. *Under the same parameters,*

$$\overline{W}^L \leqslant \overline{W} \ .$$

The inequality is strict whenever $V(\overline{W}) < L$ *and* $V^L(\overline{W}^L) = L$. *For all* $W \in [R, \overline{W}^L]$,

$$V^L(W) \leqslant V(W) ,$$

$$\overline{W}^L - W \leqslant \overline{W} - W .$$

Corollary 2.2.1 states if the principal's participation constraint binds under the limited commitment constraint, the principal can no longer defer cash payment to the agent as much as she would like to under full commitment. This also implies $V^{L\prime\prime}(\overline{W}) < 0$, that is, the "second best" frontier $rV(W)+\gamma W = \mu$ cannot be reached, suggesting a further welfare loss associated with the limited commitment contract. Since lower payment boundary implies higher likelihood of contract termination, the total surplus generated by the contract is lower when the commitment constraint binds. This provides the explanation for why firm value is always lower for any given W under the limited commitment contract: conditional on delivering the same utility to the agent, the continuation utility for the principal is lower due to earlier termination. This result is not surprising given that the self-enforcing contracting space is a strict subset of the contracting space under full commitment and $V(W)$ measures the highest value for the principal under any incentive compatible contract.

The third conclusion of Corollary 2.2.1 leads to some intuitive implications of comparing limited commitment contracts with full commitment contracts. Under the same parameter value, being closer to the payment boundary implies a higher likelihood of reaching the boundary given a certain period of time. In other words, compare two managers with the same level of continuation utility, the one under the limited commitment contract is more likely to receive bonuses in a short period of time. Meanwhile, managers under the limited commitment contract also face a higher turnover rate, because they will not be able to build a large continuation value as a result of the early payment. In all, whenever limited commitment is a binding constraint, firms' capacity to promise future payments is correlated with firm value. In normal times, firm value is high and compensation should be more "front-loaded"; whereas in crisis times, firm value is low and compensation should be more "back-loaded". Although intuitive in light of Corollary 2.2.1, these statements are so far only heuristic–I will establish them formally in the next section where I introduce the mathematical concepts needed to conduct the analysis.

Assuming limited commitment actually expands the space of parameters in which the optimal contract exists: the case when $r = \gamma$. In the full commitment model, $r = \gamma$ means the principal can costlessly delay payments to the agent. The payment boundary is therefore infinity, and the optimal contract does not exist. In contrast, the limited commitment constraint puts a physical bound on the payment boundary such that the payment boundary is where the limited commitment constraint binds.

2.3 Optimal Compensation with Uncertainty Shocks

In this section I introduce uncertainty shocks. I allow the volatility of cash flows to be stochastic, representing the transition between normal and crisis times, and derive the optimal contract under both full and limited commitment. I then show the implications of different contracts for the agent's compensation first through numerical examples from simulations and then with formal analytical arguments.

2.3.1 Volatility Regime Switching

The comparative statics above offer some intuition over the expected compensation and contract length regarding different levels of uncertainty when firms cannot commit to the contract termination time. While interesting, these comparative statics alone are not sufficient to make a compelling argument for the high level of compensation observed in crisis times. In practice, firms can make state-contingent payments. In other words, the principal can deploy the agent's continuation utility W_t optimally across different economic regimes. Comparative statics derived by holding W_t constant cannot capture the full dynamics of compensation in the presence of state transition.

To characterize the transitional dynamics of compensation under uncertainty shocks I extend the model by introducing regime switching. Specifically, I assume there are two states of the economy: σ_l and σ_h, with $\sigma_l < \sigma_h$, representing "normal" and "crisis" times respectively. Importantly, the state $s \in l, h$ is verifiable and can thus be contracted on. Given this assumption, the cash flow Y_t of the project follows

$$dY_t = \mu(e_t)dt + \sigma_s dZ_t$$

If the current state is s, in any given time interval $(t, t + dt)$, the transition probability to the other state $\widehat{s}$ is $\pi_s dt$. In the remainder of this chapter, I further simplify the model by assuming that $\pi_h = 0$, so the state h is absorbing. This means the economy starts with low volatility σ_l and experiences a one-time transition into the high volatility state with probability $\pi_h dt$ within any time interval dt. I will refer to this one-time change in volatility as the "uncertainty shock" to the economy. Although this may sound restrictive, most of the analytical results do carry through when I allow the states to be recurring, i.e. when $\pi_h > 0$. Discussion of the optimal contract under recurring states is given in the appendix[9].

Again, to offer a benchmark, assume for a moment that the principal has full commitment power. The same martingale method used to solve the single state optimal contract can be applied here but with the inclusion of an extra term in the dynamics of W_t to account for the state transition. Let N_t denote the total number

[9] I also assume that π_l is a small number to ensure that states l and h have their proper definitions. If π_l is too large, the value function (derived later) in the low volatility state converges to the value function in the high volatility state. To keep them sufficiently distant, π_l must be small enough. See Appendix A for more a detailed discussion.

of state transitions before time t. The martingale representation theorem implies

$$dW_t = \gamma W_t - dI_t - \lambda C \mathbb{1}_{\{e_t = \underline{e}\}} dt + \beta_t (dY_t - (\mu - C \mathbb{1}_{\{e_t = \underline{e}\}}) dt) + \delta(W_t)(dN_t - \pi_t dt) \quad (2.5)$$

The contract is still incentive compatible as long as $\beta_t \geqslant \lambda$. There are now two value functions for the principal $V_s(W_t)$, one for each state $s \in (l, h)$, that satisfy:

$$\begin{aligned} rV_s(W_t) &= \max_{\beta_t \geqslant \lambda, \delta_s} \mu + (\gamma W_t - \pi_t \delta_s(W_t)) V_s'(W_t) + \frac{1}{2}\beta_t^2 \sigma_s^2 V_s''(W_t) \\ &\quad + \pi_s \left(V_{\hat{s}}(W_t + \delta_s(W_t)) - V_s(W_t) \right) . \end{aligned} \quad (2.6)$$

The principal now chooses β_t and $\delta_s(W_t)$ optimally. The choice of β_t remains the same as when there is only one state since the moral hazard problem is unchanged, that is $\beta_t = \lambda$ is optimal for a incentive compatible contract. The variable $\delta_s(W_t)$ denotes the discontinuous adjustment in the agent's continuation utility at the time of regime switching. Such adjustment exists because the shadow value of the agent's continuation utility is different for the principal in different states.The principal can promise future compensation conditional on the state of the economy and substitute immediate payments with more future payments if the value of the agent's continuation utility is higher in one state. Therefore, the choice of $\delta_s(W_t)$ is determined by matching the first order derivatives of the principal's value functions before and after the regime switching, that is

$$V_{\hat{s}}'(W_t + \delta_s(W_t)) = V_s'(W_t), \text{ if } W + \delta_s > R , \quad (2.7)$$

$$\delta_s(W_t) = R - W_t, \text{ otherwise} , \quad (2.8)$$

In other words, the principal optimally deploys the agent's continuation utility until its marginal value to the principal is equalized across states. In the case where the first order derivatives cannot be matched for any δ_s that keeps the agent's continuation value in the high volatility state above his reservation utility, the contract is simply terminated.

The optimal contract still features a termination boundary R and payment boundaries $\overline{W}_s$ for each state. Boundary conditions under the full commitment setting are $V_s(R) = L$ ("value matching"), $V_s'(\overline{W}_s) = -1$ ("smooth pasting"), and $V_s''(\overline{W}_s) = 0$ ("super contact"). Payment boundaries are on the "second best" frontier $rV(W) + \gamma W = \mu$ for both the low and high volatility state. Numerical examples of the principal's value functions are illustrated in Panel A of Figure C.1. This figure shows that firm value is always lower in the high volatility state for any given level of agent's continuation utility except at the termination boundary, that is, $V_h(W) < V_l(W)$ for every $W > R$. Intuitively, since cash flow serves as a signal for the principal to infer the agent's private effort, a more noisy signal increases the likelihood of contract termination which is a necessary but costly action for the principal to provide proper incentives. That is why the regime switching from the low to the high variance state is referred to as a negative shock in this chapter.

When the principal has only limited commitment, an argument similar to Section 3.2 applies. The optimal contract features firm value functions $V_s^L(W)$ satisfying the same system of ODEs and same boundary conditions except for the "super contact" condition, which now follows the condition for the limited commitment contract proposed in Proposition 2.2.1. Specifically, for each state, if the firm value at the payment boundary is sufficiently high, then $V_s^{L\prime\prime}(\overline{W}_s^L) = 0$ is true. Otherwise, if in any state, firm value becomes too low when W approaches $\overline{W}_s^L$, then the limited commitment constraint $V_s^L(\overline{W}_s^L) = L$ binds in that state.

To summarize, the incentive compatible optimal contract under uncertainty shocks is characterized by the following proposition:

Proposition 2.3.1. *The optimal contract under volatility regime switching with full commitment defines a pair of value functions $V_s(W)$ and payment boundaries $\overline{W}_s$,*

$s \in \{l, h\}$*such that*

$$rV_s(W) = \mu + (\gamma W - \pi_s \delta_s(W)) V_s'(W) + \frac{1}{2}\lambda^2 \sigma_s^2 V_s''(W)$$
$$+\pi_s \left(V_{\hat{s}}(W + \delta_s(W)) - V_s(W)\right) , \quad (2.9)$$

subject to boundary conditions $V_s(R) = L$; $V_s'(\overline{W}_s) = -1$; *and* $V_s''(\overline{W}_s) = 0$. $\delta_s(W)$ *is determined by (2.7) and (2.8).*

If the principal has only limited commitment, the optimal contract defines a pair of value functions $V_s^L(W)$ *and payment boundaries* $\overline{W}_s^L$, $s \in \{l, h\}$, *such that* $V_s^L(W)$ *satisfies the same system of ODE (2.10) and boundary conditions* $V_s^L(R) = L$; $V_s^{L\prime}(\overline{W}_s^L) = -1$, *and*

$$V_s^{L''}(\overline{W}_s^L) = 0, \text{ if } V_s^L(\overline{W}_s^L) \geqslant L ,$$

$$V_s^L(\overline{W}_s^L) = L, \text{ otherwise} .$$

The boundary conditions specified in Proposition 2.3.1 imply that under limited commitment, the optimal contract takes three different forms depending on whether the principal's participation constraint (2.4) is binding at the payment boundary in each state: first, if (2.4) is not binding for either $s = l$ or $s = h$, this contract is simply identical to the one characterized in Proposition 2.2.1. Whether or not the principal can fully commit does not affect the contract. Secondly, the limited commitment constraint can be binding in the high volatility state but not the low volatility state. Third, the constraint can be binding in both states.[10]

Of the three types of contracts, the first type is obviously the least interesting since it is identical to the contract with full commitment. The second type can resemble contracts of either the first or the third type, depending on specific parameter values.

[10] It is impossible for the constraint to be binding in the low volatility state but not in the high volatility state, since firm value is always lower when volatility is higher. See Lemma A2.1 in the Appendix for details.

I leave the details of this type of contracts to the appendix. The third type of contract produces the most distinct implications for the dynamics of compensation between the full commitment and the limited commitment case. In the remainder of this chapter, I will concentrate discussion on this type of contract. That is, unless stated otherwise, I assume the parameter space is such that under full commitment, $V_s(\overline{W}) < L$ for both $s = l$ and $s = h$[11].

Numerical examples shown in Figure C.1 contrast contracts under different commitment assumptions given this parameter space. Recall that panel A shows the value functions for the full commitment contract. Noticing that in both states firm value is below the liquidation value L and therefore payment boundaries in neither state can be sustained without principal's full commitment. Panel B shows the value functions under the same parameters as Panel A but after imposing limited commitment.

The most crucial difference made by imposing the limited commitment condition is the position of the payment boundary. The following result highlights the point:

Corollary 2.3.1. *If $V_s(\overline{W}) < L$ and $V_s^L(\overline{W}^L) = L$ for both $s = l$ and $s = h$, then $\overline{W}_h > \overline{W}_l$; $\overline{W}_h^L < \overline{W}_l^L$. That is, the payment boundary under high volatility is higher for the full commitment contract but lower for the limited commitment contract.*

Corollary 2 states that the principal defers payments to the agent when volatility is high under the full commitment contract. Since the cost of providing incentives to the agent is the possibility of early termination after sufficiently poor performance, it is higher when volatility is higher, as rising uncertainty of cash flows increases the likelihood of sufficiently poor performance and the subsequent early termination.

[11] The exact space of parameters satisfying such condition is difficult to characterize. However, $\overline{W}$ is larger and $V(\overline{W})$ smaller whenever γ is closer to r, holding other parameters constant. This implies if the principal and the agent have similar patience level, there is a potentially large parameter space in which the limited commitment constraint will be binding in both states once it is imposed.

The principal adjusts the contract optimally by giving the agent more financial slack. Here financial slack, defined as $\overline{W}_s - R$, measures how much loss the principal is willing to take before terminating the agent's contract. Greater flexibility to the agent regarding his performance lowers the possibility of costly early termination and is thus optimal for the principal under higher volatility.

In contrast, if the principal has only limited commitment ability, payments to the agent are expedited. Limited commitment implies that the principal's participation constraint $V_s^L(W) \geqslant L$ must be satisfied at any given time. In other words, the contract must guarantee firm value of at least L, which restricts the amount of future cash flow generated by continuing the project that can be used as compensation to the agent. When uncertainty becomes higher, the total value of the project is lower. A principal lacking the ability to commit to future payments when firm value is too low is forced to pay the agent earlier because the principal can now credibly promise less compensation in the future. These relative positions of the payment boundaries under each volatility state determine the timing of the cash payment to the agent, the expected length of the contract, as well as the concavity of the principal's value function, all of which are essential in studying the compensation structure in the next section.

It is worth noting that the discontinuity in the agent's continuation utility δ has non-trivial solutions even when the transition probability π_l approaches zero, that is when the pair of value functions $V_s(W)$ converge to two independent functions with different values of variance. The effect of π_l on determining δ is small when π_l is close to zero because $V_s(W)$ moves relatively little. This implies analyses of δ can be made almost independently of π_l for small π_l which greatly simplifies the mathematics.

2.3.2 Numerical Illustration

The optimal contracts under full and limited commitment differ in how the payment boundary is determined. Their implications for compensation thus also differ, as the agent only receives payments in the form of cash bonuses once his continuation utility W exceeds the payment boundary. In this section I show how considering the optimal contract under limited commitment generates conclusions about compensation that standard contracts with full commitment cannot explain. Specifically, I argue that the large compensation observed during the recent crisis can result from optimal contracts under limited commitment, and that managers who receive large cash bonuses also face a shorter expected contract length. Contract termination can be equivalently interpreted as managerial turnover or firm liquidation in the model. I will focus on the former interpretation when discussing the results.

I begin the analysis with numerical simulations, in order to provide a transparent view of contract dynamics. In the simulations I segment the continuous-time model into discrete time intervals. The economy starts with low volatility, and the agent's initial wealth W_0 is drawn uniformly from the interval $(R, \overline{W}_l)$. I simulate N different paths of cash flows. Each path can be interpreted as one manager running an independent project. I then allow the state to switch to σ_h following a poisson arrival process, representing the transition into the crisis time. I calculate W for each of the realized cash flows and, given the payment boundaries, record the timing as well as the size of the cash bonuses. Finally, for each period before and after the uncertainty shock, I calculate the frequency of cash payments by taking the average number of recorded payments among all firms still surviving after the crisis.

I repeat this simulation procedure for both the full commitment and limited commitment contract. Results are shown in Figure C.2, with $N = 5000$.

Figure C.2 plots the frequency of payments in Panel A plots and the fraction of active projects (managers) at each given time in Panel B. Both contracts share exactly the same parameter values. They differ only in whether or not the commitment constraint is imposed, which alters their payment boundaries. I choose the parameters such that once the limited commitment constraint is imposed, it will be binding in both low and high volatility states, which most clearly manifests the implications of limited commitment.

Two observations emerge from the frequency of cash bonuses shown in Figure C.2. On the one hand, under the limited commitment contract, the frequency of payments in the few periods immediately after the uncertainty shock is much higher than the frequency under the full commitment contract and the frequency in the low volatility state. On the other hand, payment frequency under the limited commitment contract quickly diminishes to zero due to a higher rate of contract termination, while it is much more persistent under the full commitment contract. These two observations can be summarized into two theoretical predictions:

Predictions: *Managers of firms with limited commitment (1) receive more cash bonuses immediately after entering the crisis time; and (2) face a higher expected turnover rate during the crisis time, compared to managers during normal times and managers of firms with full commitment*

Both results can be formalized using mathematical concepts in stochastic calculus and are rigorously proven in the next subsection. Here I offer readers with a general interest a heuristic derivation and an intuitive explanation of the mechanism behind these results.

Frequency of cash payments can be rationalized when considered jointly with the likelihood of contract termination. When uncertainty is higher, firms with full commitment power optimally set higher bonus hurdles so managers are able to build large continuation utility, reducing the likelihood of early contract termination. In contrast, without full commitment power, large deferred payments are no longer credible. The higher the cash flow uncertainty, the lower the value from running the firm and the more likely it is for firms to terminate managers' contracts before any bonuses are realized. Managers thus have to be compensated with bonuses early for the increased likelihood of turnover. Put differently, during crisis when firm value is low, firms have to make higher payments to retain their managers, who are more worried about loosing their jobs in the near future. A similar argument applies to the comparison between normal and crisis times for the limited commitment contract, under which the capacity of credibly deferring payments is correlated with firm value in each state.

It is important to clarify here that, despite the above description of immediate payment as a result of shorter expected tenure, the two are not fruit and tree to each other but rather two sides of the same coin. Both compensation and contract length are endogenously determined by the dynamics of the state variable W. In the high volatility state, the dynamics of the agent's continuation utility are given by $dW_t = \gamma W_t - dI_t + \lambda dZ_t$. Payment dI_t reduces W_t and thus increases the likelihood of termination. As shorter length stimulates more front-loaded contracts, a front-loaded contract also implies more aggressive managerial replacement following negative performance.

Both the result regarding compensation and the result regarding turnover are supported by empirical evidence. Besides the level of compensation during the recent crisis which motivates this chapter, it has also been suggested that the high level of cash bonuses can be attributed to firm setting lower bonus hurdles[12]. The prediction that managers face higher turnover rate during market downturns is also consistent with empirical studies such as Jenter and Kanaan (2010), Kaplan and Minton (2012), Eisfeldt and Kuhnen (2013). The empirical finding of significant cor-

[12] See the Deloitte Directors' Remuneration Report (2010) and related articles on The Times and on Management Today

relation between managerial turnover and poor market returns is puzzling, given that market returns are beyond managers' direct control. Note that those studies usually focus on CEOs while this paper applies to a broader range of employees; they also do not usually control for managerial payment in their regressions and are therefore unable to directly test the theoretical hypotheses proposed here. Empirical work that simultaneously studies managerial compensation and turnover during market downturns is a potentially interesting direction for further research.

The dynamics of compensation revealed in this section also explain the variety of contracts used in practice. If investors cannot promise to refrain from withdrawing their investment when firm value drops below a certain level, managers will be hesitant to agree to a contract with *back-loaded* payments, that is, a contract that postpones most payments until satisfactory performance is reached. Notice that here satisfactory performance does not necessarily increase investors' valuation of the firm, because the firm's labor bill grows larger as well and, in the model where $V'(W_t) < 0$, better performance from the agent implies less value to the principal because the agent is paid a higher share of the profit. The manager's concern is greatest precisely during crisis, when total value from the firm's projects is the lowest, and firms are more likely to close in the near future. This leads managers to demand *front-loaded* contracts instead, where they are paid sooner rather than later as suggested by the simulation results. On the contrary, if the the principal can fully commit to retain managers they expect a longer tenure and may agree to postpone their payment further to achieve a higher total payoff from the contract.

In the aftermath of the recent financial crisis many economists and politicians blamed the current managerial compensation scheme for not aligning managerial incentives with long-term investor benefit. Consequently, policy recommendations to propagate the use of delayed payment as a solution to that problem were suggested. For instance, the Troubled Asset Relief Program (TARP) limits the ability of executives of TARP firms to cash out their restricted stock until the government is repaid in full [13]. However, the effectiveness of such recommendation hinges on the credibility that future payment promised to the executives will be delivered at full value. If executives believe that when firms are in distress, investors will withdraw by selling their shares, then the value of their stock holdings is less the longer they have to wait to cash them out. As a result executives may require even higher and more immediate compensation at the time of distress.

I also calculate the average size of cash payments for each period before and after the uncertainty shock, which produces a pattern almost identical to that observed in Figure C.2. This is not surprising given that conditional on receiving payments, the size of payments depends only on the variance σ which is a constant once the state is fixed. Moreover, instead of drawing the agent's initial utility randomly and uniformly from $(R, \overline{W}_l)$, I also conduct similar simulations but fix the manager's initial utility to be $W_l^* \equiv \arg\max_W V_l(W)$, which is the optimal level of continuation utility promised by the principal if she were to offer the contract. Results of this exercise are again very similar to those in Figure C.2 and are thus omitted here.

2.3.3 Formal Analysis of the Regime Switching Model

While numerical simulations provide intuitive and transparent stories, I now formally state and prove the results. In addition to being mathematically rigorous, the formal argument also provides new insights into some important and controversial topics in the research on executive compensation.

The argument for more immediate payments under the limited commitment contract consists of two major steps, both of which are stated relative to the full commitment contract: (1) since the payment boundary is lower when volatility is higher, at

[13] See TARP Standards for Compensation and Corporate Governance, 74 Fed. Reg. 28,394, 28,410 (June 15, 2009)

the time of regime switching, the agent's continuation utility is adjusted downward, closer to the payment boundary; (2) when the drift of the agent's continuation utility process is low, being closer to the payment boundary implies a higher likelihood of reaching that boundary and incurring cash compensation in a shorter period of time, suggesting more frequent payments immediately after the crisis hits. The argument for a higher turnover rate is relatively simpler: the agent's adjusted continuation utility after the volatility increase is also closer to the termination boundary, because the limited commitment contract gives managers less financial slack when the cash flow signal is noisier.

Adjustment in the Agent's Wealth

The first step in formally showing the result of crisis time compensation is to derive the adjustment of the agent's continuation utility $\delta(W)$. Consider the full commitment contract first. The optimal contract characterized in Proposition 2.3.1 suggests that W is discontinuous at the time of the uncertainty shock. Such discontinuity arises because the principal adjusts the shadow value of the agent's wealth optimally to keep incentives the same before and after the shock, which is reflected by the slope matching procedure introduce in Propositions 2.3.1. However, such equalization of incentive is not always viable as the two value functions corresponding to each state have different ranges of slopes. In particular at the termination boundary, because $V_s(R) = L$ for both s but $V_l(W) > V_h(W)$ for all W imply $V_l'(R) > V_h'(R)$. By concavity, $V_h'(R) > V_h'(W)$ for every $W > R$. Therefore, there is some cut-off level W_F such that if the agent's wealth before the uncertainty shock, denoted by W_{t^-}, falls below W_F, the principal simply cannot keep the same marginal value of agent's wealth after the shock. Consequently the contract is terminated as soon as the shock occurs.

This same argument applies to limited commitment contracts, but the difference lies in the sign of $\delta_l(W)$ (full commitment) and $\delta_l^L(W)$ (limited commitment) given each W. Figure C.3 Panel A illustrates the change in $\delta_l(W)$ and $\delta_l^L(W)$ as functions of the agent's wealth W before the state transition. Both the full commitment and limited commitment contracts are presented to offer comparison. Note that there is a kink point, before which $\delta_l(W)$ and $\delta_l^L(W)$ both first starts from 0 and then descends until the kink. This represents the region in which contracts are terminated once the uncertainty shock arrives. However, the full commitment and limited commitment contracts behave very differently thereafter: while $\delta_l(W)$ for the full commitment contract continues to grow until it becomes positive, $\delta_l^L(W)$ remains negative all the way to the payment boundary for the limited commitment case. This suggests that if the agent has accumulated sufficiently good performance before the shock, the size of $\delta(W)$ takes different values depending on the different payment boundaries specified according to the type of the contract.

The observation from Panel A of Figure C.3 can be formally summarized in the following proposition:

Proposition 2.3.2. *There exist cut-off levels of the agent's continuation utility* $\widehat{W}$ *such that if* $W_{t^-} > \widehat{W}$*:*

$$\delta_l(W_{t^-}) > 0$$

$$\delta_l^L(W_{t^-}) < 0$$

This proposition links the type of contract to different predictions of the "pay-for-luck" phenomenon documented by empirical works such as Bertrand and Mullainathan (2001) and Axelson and Baliga (2009). These studies find that manager's compensation is related to market performance beyond their control, which is contradictory to earlier standard contract theories such as Holmstrom (1982) which argues that the principal should filter out any signals unrelated to the manager's own performance. Many theories motivated by this contradiction have been developed in recent years, most of which feature some kind of managerial entrenchment or hidden talent that is partly related to market signals. See Bebchuk and Fried (2006) and Jenter and Kanaan (2010) for a more detailed survey.

In contrast to existing theories, pay-for-luck is a natural feature of the regime switching model used in this chapter. As the agent is not responsible for the occurrence of uncertainty shocks, his continuation utility from the contract is indeed adjusted to keep its marginal value to the principal unchanged. However, whether that adjustment corresponds to a "reward" or a "punishment" depends on the principal's commitment power. The full commitment contract rewards the agent with higher continuation value when uncertainty is higher. Considering higher uncertainty naturally as "bad luck" since firm value is lower in this state, this prediction is less intuitive. On the contrary, when the principal has limited commitment, agents are "punished" for higher uncertainty as their continuation value is brought down. In fact, if the model allows recurring state transitions, the direction of δ flips signs when the state switches from "crisis" to "normal", and the limited commitment contract predicts a "reward" for "good luck", which is largely consistent with the empirical findings.

It should be noted nonetheless that here neither the "reward" nor "punishment" involves any instant cash transfer. As W measures the agent's present value of all future payments, the adjustment of W is merely a reflection of the different payment boundaries and termination likelihood. The actual payments are related not only to the shift in boundaries but also the distance between the boundaries and the agent's continuation value after the adjustment. The next result based on Proposition 2.3.2 sheds light on this point:

Corollary 2.3.2. *Let W_{t^-} be the agent's continuation utility before the uncertainty shock, and $W_{t^+} \equiv W_{t^-} + \delta_l(W_{t^-})$ and $W^L_{t^+} \equiv W_{t^-} + \delta^L_l(W_{t^-})$ be the agent's continua-*

tion utility after the uncertainty shock under full and limited commitment contract, respectively, then

$$\overline{W}_h^L - W_{t^+}^L < \overline{W}_h - W_{t^+}, \textit{ if } W_{t^-} > \widehat{W} \; .$$

This result is illustrated in Panel B of Figure C.3. It shows that while the payment boundary of the limited commitment contract is lower, the agent's adjusted continuation utility after the uncertainty shock is also closer to the payment boundary. This conclusion plays a leading role in the analysis of compensation later, as being close to the payment boundary implies a larger probability of receiving more cash payments in the near future. At the same time, a lower payment boundary suggests a higher likelihood of contract termination following a series of poor performances. This trade-off between immediate cash payments and likelihood of termination is the central mechanism behind the dynamics of compensation.

After establishing the direction of $\delta(W)$ and the position of W_{t^+} relative to the payment boundary, I can formalize the observations from the simulation example using standard methods in stochastic calculus. The argument is presented in the next subsection

Analytical Characterization

Here I characterize the dynamics of compensation following uncertainty shocks. Given any W_{t^+}, the agent's continuation utility after the volatility increase, the goal is to characterize the distribution of the agent's wealth after a certain amount of time elapses. Following Cox and Miller (1977), given the dynamics of W, the transition density function $f(t, W; W_{t^+})$ for a process starts with W_{t^+} and satisfies the Kolmogorov forward equation:

$$\frac{\partial}{\partial t} f(t, W; W_{t^+}) = \frac{1}{2} \frac{\partial^2}{\partial W^2} \left[\lambda^2 \sigma_h^2 f(t, W; W_{t^+}) \right] - \frac{\partial}{\partial W} \left[\gamma W f(t, W; W_{t^+}) \right]$$

subject to boundary conditions:

$$f(t, R; W_{t^+}) = 0$$

$$\frac{1}{2}\frac{\partial}{\partial W}\left[\lambda^2\sigma_h^2 f(t, W; W_{t^+})\right]|_{W=\overline{W}_h} - \gamma\overline{W}_h f(t, \overline{W}_h; W_{t^+}) = 0$$

Unfortunately, this partial differential equation is generally intractable. However, when γ is small, the dynamics of W can be approximated by a standard Brownian motion with one absorbing boundary R and one reflecting boundary $\overline{W}_h$, whose transition density has an explicit form[14]. Details on the approximation and the derivation of the transition density are shown in the Appendix by virtue of the method developed in Ward and Glynn (2003)

After obtaining the transition density, I can measure the likelihood of cash payments given a certain time period T after the shock using the concept of local time in stochastic processes. Given a time period T and initial point W_{t^+}, define local time $\mathcal{L}$

$$\mathcal{L}_h(T; W_{t^+}) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T \mathbb{1}_{\left\{\overline{W}_h - \varepsilon < W_t < \overline{W}_h + \varepsilon\right\}} dt | W_0 = W_{t^+}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function. This local time is a random variable that measures the amount of time W spends in the neighborhood of the payment boundary. Since being at the payment boundary implies cash payments, this can be interpreted as the frequency of payments an agent with initial wealth W_{t^+} receives within time T after the economy enters crisis mode.

The interesting value is the expectation of local time given by

$$E\left[\mathcal{L}_h(T; W_{t^+})\right] = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T dt \int_{\overline{W}_h - \varepsilon}^{\overline{W}_h + \varepsilon} f(t, W; W_{t^+}) dW$$

[14] The assumption $\gamma > r$ is still needed for the benchmark full commitment contract to exist. It is not necessary, though, for the limited commitment contract. See the end of Section 3.2. for the discussion

This value measures the expected frequency of cash payments given initial wealth W_{t^+}. The higher this value, the more frequently the agent can expect to receive cash payments before time T during the crisis. As the numerical simulations show, under the limited commitment contract, agents on average receive cash payments more frequently for a short period immediately after the crisis begins. This observation can now be stated formally using the expected local time defined above. Last but not least, the numerical simulations begin with a random agent's wealth in the low variance state. In order for the analytical results to better match those from simulations as well as what is seen in practice, I need to replace the fixed initial wealth W_{t^+} in the high variance state with the agent's wealth W_{t^-} in the low variance state. Thanks to Proposition 2.3.2 and Corollary 2.3.2 there is a one to one mapping between the two variables which allows me to summarize the analytical finding in the next proposition.

Proposition 2.3.3. *Assume γ is small. There exists $\widehat{T}$ and $\widehat{W}_{t^-}$ such that if $T < \widehat{T}$:*

$$E^L\left[\mathcal{L}_h(T; W_{t^-})\right] > E\left[\mathcal{L}_h(T; W_{t^-})\right]$$

$$E^L\left[\mathcal{L}_h(T; W_{t^-})\right] > E^L\left[\mathcal{L}_l(T; W_{t^-})\right]$$

for all $W_{t^-} > \widehat{W}_{t^-}$, where E^L represents expectation under the limited commitment contract

Proposition 2.3.3 provides the formal conditions under which the frequency of payments is higher under the limited commitment contract, in the high volatility state. Despite the mathematical complexity, its basic intuition is quite simple: first, compare the limited and full commitment contract, Corollary 2.3.2 shows W is closer to the payment boundary after the uncertainty shock under the limited commitment contract. When γ is small, the process of W behaves similarly to a standard Brownian motion and thus spends more time at the payment boundary whenever the starting point is closer to the boundary. In more intuitive terms, the agent should expect more frequent payments in the near future if his cumulative performance is closer to the target bonus hurdle set by his contract. The similar argument applies to the comparison between the limited commitment contract in low and high volatility states: W_{t^+} is closer to $\overline{W}_h^L$ than W_{t^-} is to $\overline{W}_l^L$.

Why does Proposition 2.3.3 hold only when T is small? This is because while W_{t^+} is closer to the payment boundary after the shock under the full commitment contract, it is also closer to the termination boundary because the agent is overall punished according to Proposition 2.3.2. As T increases, the likelihood of contract termination rises faster for the limited commitment contract. That is, agents now operate under tighter financial slack. The longer into a crisis, the more likely is termination, as the possibility of realizing a series of losses becomes more real. The conclusion in Proposition 2.3.3 thus holds only for T small enough, when the probability of termination is negligible. As shown by the numerical simulations, this pertains to the second observation that cash payment vanishes very quickly under the high volatility state under limited commitment. The notion of termination likelihood can be formally described using the concept of stopping time, as the next proposition shows:

Proposition 2.3.4. *Define* $\tau_s = \inf\left\{t : W_t = R|\overline{W}_s\right\}$ *as the termination time given payment threshold* $\overline{W}_s$ *then:*

$$E^L(\tau_h) < E^L(\tau_l)$$

$$E^L(\tau_h) < E(\tau_h)$$

When the commitment constraint is binding, the agent's expected termination time is shorter under high volatility

Intuitively, given the absorbing boundary R and reflecting boundary $\overline{W}_h$, a process with initial value W_{t^+} is in expectation stopped earlier whenever W_{t^+} is closer to R and $\overline{W}_h$ is smaller. The limited commitment contract satisfies both conditions. Further, it should be noted that this proposition does not require the assumption of a small γ, as the expected speed of growth for W is lower when W_{t^+} is lower, which the limited commitment contract again satisfies. Nevertheless the proof of Proposition 2.3.4 still imposes the restriction on γ for the sole purpose of analytical tractability.

The results of this subsection imply that the recipients of crisis time bonuses are those who perform relatively well before the crisis. Proposition 4 states that more frequent cash compensation is conditional on the agent's wealth before the shock W_{t^-} surpassing a certain threshold, and higher W_{t^-} represents better before-shock performance. Those who perform relatively poorly ex ante are no longer around after the crises as a result of either replacement or firm liquidation. Combined with Proposition 2.3.2, this suggests that those who produce the largest profits before the crisis are being criticized the most for receiving bonuses during the crisis. One should keep in mind, however, that the huge loss of firm wealth is primarily due to the risky aggregate environment and, despite receiving bonuses for a short period into the crisis, managers are being harmed overall.

The optimal contract derived in this chapter is not renegotiation-proof, which may raise a legitimate concern but does not affect the main results. Renegotiation-proof contracts require the value function to be downward sloping everywhere. In the Appendix I derive the renegation-proof contract and show that the main results carry through. Despite the principal having only limited power of commitment, renegotiation-proofness is not a necessary feature of the resulting equilibrium contract, because it is assumed in the model that the principal can commit not to renegotiate the contract but just cannot commit to when to terminate the contracting relationship.

The assumption of constant liquidation value L and reservation utility R for both high and low volatility states is for the sake of simplicity but could be extended to better match reality. It is quite plausible that both outside options could have state-contingent values. Managers could have difficulty finding another job if laid off during crisis times. Similarly, if L is interpreted as the fixed cost for firms to replace an incumbent manager, both L and R should be lower in the high volatility state.

Nevertheless, allowing reservation value to be state-contingent does not affect the validity of the main results. The renegotiation-proof contracts discussed in the Appendix feature endogenous renegotiation boundaries, which are lower when volatility is higher, similar to assuming a lower R in the high volatility state. The main results of this chapter still hold under the renegotiation-proof contracts because they depend mainly on the relative positions of the payment boundaries. As for the liquidation value L, a lower value implies a lower value for the principal's outside option, thus reducing the tightness of the commitment constraint imposed on the contract. Therefore, whether more frequent cash bonuses are paid in the high volatility state simply depends on the decrease in L relative to the increase in σ.

2.4 Extensions

In this section I discuss two extensions of my main model. One, the implementation of the limited commitment contract, which justifies the limited commitment constraint by revealing a similarity between a firm's commitment to a contract termination time and its commitment to a capital structure. Two, I explore the equilibrium in which shirking is optimal and its implications for both empirical studies and policy recommendations

2.4.1 Contract Implementation, Capital Structure and the Commitment Constraint

Results from the previous section highlight different dynamics of compensation generated by full commitment contracts and limited commitment contracts. In this section I further explore the difference between those contracts via contract implementation and establish a novel equivalence between firms' commitment to compensation contracts and commitment to capital structure. The equivalence provides a justification for the limited commitment assumption made throughout this chapter, as firms' commitment to capital structure is known to be implausible.

Implementing the full commitment contract involves the use of debt and equity, and thus creating a conflict between debt and equity holders that leads to potential commitment issues. When the regime switches from low to high volatility, the face value of debt must be brought down at the expense of equity holders. Such implementation imposes an implicit assumption that equity holders must commit to maintain a certain capital structure, which is generally implausible since equity holders do not always act for the benefit of the entire firm. In contrast, contracts with limited commitment do not require firms to make such commitment to capital structure and should thus be more prevalent in practice.

The implementation in this chapter follows the standard literature in using a set of common securities with limited liability: equity, long-term debt, and credit line. Equity can be held by both the manager as well as outside investors who receive dividend payments and can decide the firm's capital structure. The long-term debt is a callable consol bond that pays a fixed rate and has a fixed face value. The firm can issue more long-term debt or call it back at its face value [15]. Finally, the credit line provides the manager with limited liquidity. The manager decides both the dividend and the credit line balance, but incentive compatibility under the optimal contract renders irrelevant who makes dividend and credit line decisions.

There is more than one implementation of the optimal contract. The following proposition provides a standard result:

[15] Although callable debt is usually redeemed at a premium, the specific value of the premium does not play a role in this model and is thus without loss of generality assumed to be zero.

Proposition 2.4.1. *Both the full and limited commitment contract can be implemented by*

(a) manager holding inside equity share λ*;*

(b) face value of the callable debt satisfying $D_s = V_s(\overline{W}_s)$*;*

(c) credit line balance M_t *and credit limit* C^* *satisfying* $W_t = \lambda(C_s^* - M_t)$ *and* $\lambda C_s^* = \overline{W}_s$.

Dividend is paid when $M_t = 0$*. Liquidation occurs when* M_t *reaches* C_s^*.

The implementation is intuitive and hence the explanation here concise. Since λ measures the portion of private benefits the manager can derive from shirking, its value represents the least degree of sensitivity to cash flow to which the manager is exposed. Managers can draw down the line of credit for operating liquidity. Dividends serve as reward to the manager as well as returns to outside investors. Since the manager has a higher discount rate, dividends will not be saved inside the firm as long as the credit line is paid in full. Finally, the amount of long-term debt is used to adjust the profit rate of the firm such that incentives remain unchanged.

Security implementation of the optimal contract implies a certain capital structure which can potentially raise questions under the regime switching environment: the boundary conditions of the full commitment contract implies $V_h(\overline{W}_h) < V_l(\overline{W}_l)$. Since $V_h(\overline{W}_h)$ and $V_l(\overline{W}_l)$ also correspond to the face value of the callable debt in the high and low volatility state, the implementation of the full commitment contract requires that the face value of long-term debt be brought down when volatility increases. In other words, some portion of the long-term debt must be called back, hence the usage of callable debt here. These callbacks induce a transfer of wealth out of equity holders' pockets while debt holders are paid in full.

To further investigate this problem, I compute the value of the aforementioned securities, in particular that of equity. To simplify the analysis, I assume that $L = 0$, so that there will be no residual value in the event of firm liquidation, eliminating the need to specify the priority of residual claims among equity holders and debt holders. Let V_s^E denote the equity value, which is defined by

$$V^E(W_t) = E\left(\int_t^\tau e^{-rs} dDiv_s | W_t\right)$$

where W_t, the manager's continuation utility, can be transferred to M_t, the credit line balance, through the relationship defined in Proposition 2.4.1.

The value function of equity value can be characterized by the following differential equation:

$$rV_s^E(W) = (\gamma W - \pi_s \delta_s(W))\, V_s^{E\prime} + \frac{1}{2}\lambda^2 \sigma_s^2 V_s^{E\prime\prime} + \pi_s(V_{\hat{s}}^E(W + \delta_s(W)) - V_s^E(W))$$

subject to boundary conditions

$$V_s^E(0) = 0$$

$$V_s^{E\prime}(\overline{W}_s) = 1$$

where $V_{\hat{s}}^E$ is the value of equity in state $\hat{s}$ The implementation requires equity holders to commit to the particular capital structure specified in the optimal contract by redeeming the outstanding debt at the time of the uncertainty shock.

Do equity holders always find it preferable to recall debt when uncertainty is high? The answer is hardly yes, as equity holders can usually withdraw investment in practice and default on any debt obligation. In this model, let $D_{\hat{s}} - D_s$ measure the value of debt redemption. Equity holders will find it optimal to default when $V_{\hat{s}}^E$, the value from maintaining the firm, is lower than $(D_{\hat{s}} - D_s)$, the cost of doing so. On the contrary, under the limited commitment contract, $V_s(\overline{W}) = L$ for both $s = l$ and h implies an identical face value of long-term debt before and after regime switching. That is, the capital structure of the limited commitment contract can be maintained without a tendency on the part of equity holders to default ex post.

Equity holders' making ex post default decisions is common in both financial research and in practice. There is a large body of literature studying the endogenous default decision of equity holders and the conflict with debt holders, notably Leland (1994), Leland and Toft (1996), He and Xiong (2012). In this chapter, I do not characterize the exact default boundary of equity holders; rather, the point I want to make is that equity holders cannot (credibly) commit to not defaulting for the sake of the entire firm when there is a chance of their finding default preferable ex post.

In additional to justifying the prevalence of limited commitment contracts in practice, the equivalence between the commitment to contract termination time and the commitment to capital structure also offers an empirically testable hypothesis: investors of more distressed firms are more likely to withdraw their investment, default on firms' debt and alter firms' capital structure. The degree of distress can be a potential proxy for the commitment power firms have over their labor contracts which is difficult to observe.

2.4.2 Optimal Compensation with Shirking

Throughout previous analyses, it has been assumed that working is always preferred by the principal regardless of the level of uncertainty. This section relaxes this assumption and examines when the optimal contract allows shirking in equilibrium. The results carry both policy and empirical implications. When the contract allows shirking during crisis times, no bonuses are paid. This may appear agreeable to policymakers and to public sentiment, but it is actually worse because the average productivity of the economy is lower as a result of lower managerial effort. Empirically, studies of compensation and performance, such as those examining pay-performance sensitivity, could be confounded by the endogeneity between return and volatility driven by unobservable changes in managerial effort.

Which effort level is optimal in equilibrium depends on the cost of allowing shirking. Working is preferred as long as C, the social cost of shirking measured by the reduction in average cash flow, is high. This section explores this assumption in more detail. Consider a contract that involves no payment but simply allows the agent to shirk forever. Define (W^S, V^S) as the pair of payoffs for the agent and the principal respectively if the agent exerts effort $e_t = \underline{e}$ for all t, then

$$W^S = \frac{\lambda C}{\gamma} \; ;$$

$$V^S = \frac{1}{r}\left(\mu - \frac{\lambda C}{\gamma}\right) \; . \tag{2.10}$$

Notice that this payoff is a function of A, which is an irrelevant variable in the incentive compatible contract characterized in Proposition 2.2.1 and 2.3.2. Therefore, whether the incentive compatible contract is optimal for the principal depends on the level of V^S. When C is sufficiently low, $V^S > V(W^*)$, where $V(W^*) \equiv \max V(W)$ is the maximal value the principal can derive from an incentive compatible contract, the principal is better off stopping incentive provision[16]. The agent will choose to shirk, receive no payment from the principal and instead be compensated by his private benefit from shirking. The optimal contract is static, unrelated to the agent's performance, and therefore involves no termination.

The different maximal firm value under low and high volatility states raises the possibility that working is not always optimal for both states. If $V_l^* > V^S > V_h^*$, the optimal contract will induce working as long as $s = l$ and switches to the static contract at times when $s = h$. In the model where only one state transition occurs, the dynamics of the optimal contract follow the ODEs described in Proposition 2.3.2, except the value function V_h is replaced by the static payoff given by equation (2.10).

[16] Strictly speaking, the contract that allows shirking forever is optimal only when $V^S > B(W)$, where $B(W)$ is a V-shaped function that extends above V_h^*. See Zhu (2012) for the details. Here I avoid the complicated situations where V^S lies above b^* but below $B(W)$ by assuming that C is either high enough or low enough so that either working or shirking permanently is the optimal effort.

If shirking is optimal in the high volatility state, the procedure that pins down $\delta_l(W)$ is slightly different. Under the static contract that allows shirking when $s = h$, the agent's continuation utility is a singleton W^S. Jumps of W from the low to the high volatility state is simply $\delta_l(W) = W^S - W$. Note that W^S measures not the discounted future income but the present value of private shirking benefit to the agent. This value is automatically achieved as long as the principal immediately ceases any payment. Moreover, because W^S is no longer sensitive to the agent's performance, there is no contract termination after the regime switches to high volatility. All firms survive regardless of their agent's performance history up to the regime switching time. The next proposition summarizes these findings:

Proposition 2.4.2. *Suppose C is low or σ_h is sufficiently high, the optimal contract induces $e_t = \overline{e}$ under σ_l but $e_t = \underline{e}$ under σ_h. The principal's value function $V_l(W)$ and payment boundaries $\overline{W}_l$ satisfy:*

$$\begin{aligned} rV_l(W) &= \mu + (\gamma W - \pi_l \delta_l(W))\, V_l'(W) + \frac{1}{2}\lambda^2 \sigma_L^2 V_l''(W) \\ &\quad + \pi_l \left(V_h - V_l(W)\right) \end{aligned}$$

subject to boundary conditions $V_l(R) = L$; $V_l'(\overline{W}_l) = -1$; and $V_l''(\overline{W}_l) = 0$. Furthermore, $\delta_l(W) = W^S - W = \frac{\lambda C}{\gamma} - W$, and V_h is given by

$$V_h = V^S = \frac{1}{r}\left(\mu - \frac{\lambda C}{\gamma}\right) .$$

The existence of an optimal contract that involves shirking in the equilibrium has important policy implications. Since the manager is compensated through the private benefit of shirking when uncertainty is high, no cash payment is made under that regime. This implies the possibility of observing little or no bonuses during a recession. However, though much to the media or public's liking, this equilibrium is actually worse in terms of total welfare, because productivity, measured by mean cash flow, is now lower due to less effort from managers. This is true as long as $\lambda < 1$ so that there is deadweight loss associated with managerial shirking. This result highlights the importance of compensation in keeping mangers properly incentivized, even though the exact timing of their compensation may not match their overall performance at the time when a large negative shock occurs.

The shirking equilibrium also reveals a potential endogeneity problem between profitability and volatility. Existing empirical work that studies compensation often considers profitability and volatility as independent factors. However, fluctuation in profitability can be driven by the change in volatility through the channel of managerial effort, raising empirical challenges since such effort is normally difficult to measure. It also provides further evidence in addition to previous work that uncertainty is the key to understanding the recent financial crisis.

It is worth noting that the change in the agent's equilibrium effort is a feature of increasing volatility but not necessarily of decreasing profitability. This sets this paper apart from those with similar regime switching techniques such as Hoffmann and Pfeil (2010). While lower average cash flow μ does bring down firm value under an incentive compatible contract, it also lowers V^S, firm value under a static contract that allows shirking. As a result, working can still be the optimal effort to induce if $V^S < V_h^*$. In contrast, V^S does not depend on σ, but V_h^* does. When cash flow volatility becomes higher, V_h^* becomes lower until falling below V^S, and the incentive compatible contract is dominated by the static contract, a unique feature of stochastic volatility.

2.5 Remarks

This chapter studies the optimal compensation contract under the twin assumptions of limited commitment by the principal and regime switching of cash flow volatility. Sudden and dramatic increases in market uncertainty have been argued as the most critical aspect of financial crises. When uncertainty is high, investment becomes more risky, and the value of continuing the firm is correspondingly low. Principals without full commitment cannot credibly pledge sufficient amounts of future payments and must provide agents with more immediate compensation. This offers an explanation for the highly controversial large compensation–especially cash bonuses–paid by many financial firms during the recent crisis. At the same time, managers are subject to a higher turnover rate. Therefore, despite the bonuses, managers are still worse off in crisis times as their overall present value from the contract is lower.

The model provides a new perspective on the current compensation structure not only in regard to bonuses. While equity-based compensation contracts provide a solution to the problem of aligning managers' incentives with those of investors, their effectiveness depends on the ability of both parties to commit to the contract. Furthermore, payment and expected tenure are two counter weights both endogenously determined in the optimal contract. Any measure that intends to provide better long-term incentives to managers must take into account both payment and the expected contract length.

There are several directions in which this model can be fruitfully extended. In the model there is only one representative firm and one representative agent, so the uncertainty shock can be interpreted as either aggregate or idiosyncratic. A model that allows firm heterogeneity and differentiates idiosyncratic shocks from aggregate shocks may generate interesting results, such as the cross-sectional predictions regarding the response of compensation to firm level investment risks. This model also potentially speaks to the the important issue of liquidity management in response to market downturns. Recent studies such as Campello et al. (2011) examine cross-sectional liquidity management along different dimensions. A slightly modified dynamic model à la Bolton et al. (2011) would be readily equipped to provide theoretical insights for these observations.

3

Dynamic Investment and Asset Prices

In the past two decades, the world has witnessed the rapid integration of international financial markets. Prior to the recent financial crisis, gross inter-country capital flows surged from 1.5 trillion US dollars in 1995 to about 11.8 trillion dollars in 2007. During this process, developed economies, such as. the US, ran large deficits whereas emerging economies, such as. China, large surpluses. Existing research on financial integration has focused mainly on its impact on the US market. However, given their explosive growth and increasing importance in the international economy, emerging markets warrant greater attention from researchers.

This chapter expands our understanding of financial integration by examining for the first time its impact on emerging markets. More specifically, this chapter studies the two most prominent effects of financial integration observed for emerging markets: (1) savings gluts, referring to the upsurge in domestic savings, and (2) asset price booms. Figure 1 compares the gross savings rate of the US to that of emerging East Asian countries and China in particular. We observe stark differences in the savings behavior between the US and the latter two. The term ”savings glut” was first addressed by Bernanke (2005), who argues that high domestic savings of many emerging economies is the source of global financial account imbalance. Subsequent research provides detailed measurement of savings gluts for cross-country and time series samples[1]. More interestingly, savings gluts in emerging economies largely take the form of foreign risk-free assets such as US treasury bills. Figure 2 shows an example of the price boom: the housing price in China. In general empirical studies have document significant asset price increase in capital outflow countries[2]. Although most noticeable in the real estate market, price booms are generally observable in other markets, such as the stock market, as well. Glindro et al. (2007), Gochoco-Bautista (2008), Igan and Jin (2010) etc offer more detailed documentation on how asset price booms are related to a country's saving behaviors.

The simultaneous occurrence of savings gluts and asset price booms implies a complementary relationship between domestic risky assets and foreign risk free assets for emerging economies. This is puzzling because the two assets are more naturally thought of as substitutes. Indeed, the current literature, which focuses on the US and generally takes as given the behavior of emerging economies, predicts substitution: increase in demand for foreign risk free assets decreases demand for and hence the price of domestic risky assets.

In this chapter I provide a highly tractable continuous time dynamic model that explains the phenomenon of simultaneous savings gluts and asset price booms observed in emerging economies during the integration of financial markets worldwide. Financial integration provides firms in emerging economies with a foreign risk free asset when they could only hold a risky domestic asset before. While firms substitute their investment in the domestic risky asset with the foreign risk free asset, i.e. the substitution effect, the introduction of the risk free asset also generates another effec-

[1] For instance Bernanke (2007), IMF (2007), Bayoumi et al. (2010), Ma and Yi (2010), Yang et al. (2010) etc

[2] See, for example, Rajan (2006a), Ferguson and Schularick (2007), Wu et al. (2012)

t: the *wealth effect*: holding a less risky portfolio because of the addition of the risk free asset, firms are more likely to accumulate large wealth before incurring losses from their investment in the risky asset. The stationary distribution of firm wealth skews more to the left in equilibrium, meaning that more firms are made richer after financial integration. When the wealth effect dominates, the overall value invested in both the domestic risky asset and the foreign risk free asset increases, leading to a higher domestic asset price as well as a higher aggregate savings rate.

There has been a number of studies analyzing the cause and consequences of such direction of capital flows. Most notably Caballero et al. (2008), Caballero and Krishnamurthy (2009), Mendoza et al. (2009), Angeletos and Panousi (2011), and Maggiori (2012). The closest to this chapter is Mendoza et al. (2009), who build a two-country model where the domestic environment of contract enforceability predicts the change of emerging countries' portfolio from their domestic risky assets to risk free assets supplied by developed countries, but the subsequent substitution effect drives asset prices in emerging countries down instead of up. Existing research, due to a US focus, has been taking emerging economy characteristics as exogenous and failed to explain why in emerging economies domestic risky assets and foreign risk free assets appear to be complements rather than substitutes.

This chapter addresses this complementarity puzzle through a continuous-time, reduced-form investment model similar to Bolton et al. (2011). Firms make payout and portfolio choice decisions but are subject to costly refinancing when their capital level is low. Because of refinancing costs, firm are effectively risk averse even though they value dividends in a risk-neutral manner. Moreover, firms' degree of risk aversion is endogenous and depends on the overall risk of their investment portfolio. Modeling in continuous time gives me the crucial advantage of analytically tracking the distribution of firm wealth, a technique widely used in the growing continuous-time macro finance studies such as Brunnermeier and Sannikov (2012), He and Krishnamurthy (2013a,b) etc. Different from those studies, I focus on the steady state analysis to simplify the solution, but this model can be easily extended to study the full dynamics of firm wealth distribution given the continuous-time method.

Considering savings gluts and asset price booms together allows new insights into understanding the impact of financial integration on emerging economies. It leads to the counterintuitive revelation that high domestic asset price and savings glut concentrated in foreign assets are fundamentally linked and reinforces each other. It also shows how different economic fundamentals, such as investment risk and return, contribute to movements in asset price and savings rate during financial integration. Comparative statics predict all of the following to magnify asset price booms and savings gluts: high domestic investment return, low domestic investment risk, low international interest rates, and high external financing costs. These predictions are broadly consistent with empirical findings on emerging markets and able to nest the conclusions of several existing theories, such as "liquidity gluts" and "investment slumps", that explain the movement in either asset price or savings rate in emerging countries.

The model builds on two important market frictions prevalent among emerging economies: first, limited commitment prevents firms from borrowing as they cannot credibly pledge future income. The combination of risky investment in domestic assets and limited commitment gives firms the demand for precautionary savings, which they meet using the risk free asset available after financial integration. Limited commitment continues to apply after financial integration such that firms still cannot borrow but are able to hold positive positions in the risk free asset. Second, costly external financing makes firms risk averse because they want to avoid the cost of raising new capital when their net worth is low. The cost of financing must be sufficiently high to result in asset price booms and savings gluts in this model. This is because higher financing costs imply stronger risk aversion and consequently a stronger wealth effect.

This chapter also broadens the implication of limited contract commitment on financial markets. Limited commitment refers to the lack of ability to credibly pledge future income when firms are short on liquidity, which is the key financial friction modeled in this chapter. Previous work by Kiyotaki and Moore (1997), Krishnamurthy (2003), Cooley et al. (2004), and Lorenzoni (2008) explores the aggregate implication under limited commitment especially for asset prices in closed economies, while this chapter focuses on an open economy. I also conduct welfare analysis in a similar manner to Lorenzoni (2008), who derives ex ante over-investment as a result of limited commitment and decentralized trading. My results share similar features but have different implications for the optimal level of asset price and the distribution of firm wealth.

The high level of asset price and savings rate resulting from financial integration is inefficient, however, compared to the socially optimal level. In the decentralized economy, firms' portfolio choice determines the equilibrium domestic asset price, imposing a pecuniary externality which firms do not internalize. A social planner, taking into account the effect of firm choices on the equilibrium asset prices, places more weight on the international risk free bond, resulting in a lower asset price and savings rate. The portfolio choice of the social planner provides a rationale for the large holdings of US treasury securities by the government of many emerging economies. Finally, the socially optimal portfolio choice can be implemented for the competitive market via government taxation and subsidy, which leads to several policy recommendations.

Savings gluts and asset price booms can be amplified and empirically predicted by heterogeneous productivity, a feature prevalent on emerging economies' markets. Recent literature has conflicting views on the role of productivity heterogeneity in determining asset prices and investment in emerging economies. I find that the larger the mean-preserving spread in firm productivity, the stronger the effect of financial frictions on asset price booms and savings gluts. This is because when firms are heterogeneous, their portfolio choice is not only affected by asset returns but also their relative risk aversion from holding risky assets. I find that firms with lower productivity are more risk averse. In autarky, this implies low asset price and savings rate because assets are less valuable. In financial integration, low productivity firms engage in more precautionary savings and consequently increase asset prices more. The results suggest the empirical implication that cross-country differences in asset price and savings rate can be explained by heterogeneity among domestic firm and the productivity level of the least productive firms.

This chapter is organized as follows. Section 3.1 describes the model and introduces the key market frictions. Section 3.2 solves and compares asset prices and savings rates under different equilibriums and presents comparative statics. In Section 3.3 I study the optimal asset portfolio choice problem from the perspective of the social planner's and make policy recommendations. In Section 3.4 I introduce productivity heterogeneity and analyze its impact on asset allocation and total productivity. Section 3.5 concludes. All proofs are in the Appendix.

3.1 Model

This section describes the basic environment of the model: one country, continuous time, infinite horizon. I define two equilibriums: *autarky*, where the only asset available for investment is a domestic asset with fixed supply and risky return; *financial integration*, where a risk free asset of infinite foreign supply is added.

3.1.1 Autarky

There is one economy representing an emerging country, whose only domestic asset is land. Land produces rice, which is perishable and used as the numeraire. Land does not depreciate nor can it be re-produced, so its aggregate quantity is fixed over time and is normalized to one unit.

The economy is populated with two types of agents: firms and households. Firms are risk neutral with discount rate ρ. They make investment in land and produce rice which can either be used to purchase more land or be paid out as dividends. A firm holding K_t units of land can produce dY_t units of rice according to

$$dY_t = K_t dA_t \ ,$$

where $dA_t = \mu dt + \sigma dZ_t$ is the *idiosyncratic* productivity of each firm. Importantly, the technology shocks, represented by the standard Brownian motion term Z_t, are assumed to be i.i.d. across all firms. In other words, there is no aggregate shock in this economy. Let dC_t denote the dividend payout during a unit of time. Firms' objective is to maximize the discounted value of total future dividends:

$$\max \int_0^\infty e^{-\rho t} dC_t \ .$$

The price of land is denoted by P_t. Firms all take P_t as given. The unit return from investing in land is thus defined by

$$dR_t = \frac{dA_t + dP_t}{P_t} \ .$$

Firms are owned by households, who consume the dividends paid by firms but cannot directly invest in land. To keep the households' problem simple, I assume they are hand-to-mouth consumers, holding completely diversified shares of all firms.

Diversification implies that households are immune to firms' idiosyncratic productivity shocks and simply consume a constant total output which equals μ. This ensures that households play a minimal role in this economy so subsequent analyses can focus on firms' decisions[3].

The key market friction in this emerging economy is limited commitment. Each firm can default on its entire debt obligation while retaining all the land and rice they have and cannot be excluded from future trading. Rampini and Viswanathan (2013b) show that such limited commitment constraint is equivalent to a borrowing capacity constraint, which in this model equals zero. As a result, there is no domestic credit market because no firm can credibly pledge any future payment; land is the only asset in which firms can invest; the price of land P_t is the only price that needs to be determined endogenously in equilibrium, which greatly simplifies the solution.

In practice, limited commitment is prevalent among emerging economies. Cooley et al. (2004) provide a measurement for contract enforceability for a large cross-country panel. Unsurprisingly, countries with less developed financial markets are generally associated with poorer contract commitment. Allen et al. (2005), Lu and Tao (2009), Du et al. (2012) document weak contract commitment particularly in China—the best example for the simultaneous savings gluts and asset price booms discussed in this chapter. Here I make the assumption of zero credible future debt payment, which completely shuts down the domestic credit market. Albeit extreme, such assumption allows a much simpler, closed-form solution of firm dynamics and the market equilibrium.

Given all the assumptions above, I now characterize the dynamics of firm net worth, denoted by W_t, which follows a stochastic process between two boundaries:

[3] I can also define households as standard risk averse utility optimizers and formally describe and solve their optimization problem. However, as long as households cannot invest directly in land but are allowed to hold completely diversified shares of all firms, the absence of aggregate risk implies perfect consumption smoothing across time, which is equivalent to what is described here.

payout boundary $\overline{W}$ and refinancing boundary $\underline{W}$. First, firms will distribute dividends only when W_t is sufficiently high, due to the risk of holding land. $\overline{W}$ is a reflecting boundary such that whenever W_t exceeds, lump sum dividends $W_t - \overline{W}$ are paid so that the process W_t immediately reflects back to $\overline{W}$. Second, firms cannot have negative net worth due to limited commitment. Because dA_t follows an arithmetic Brownian motion, large losses can occur and firms' net worth approaches zero, at which time firms must refinance by raising more capital[4].

The limited commitment assumption also has important implications for how firms can refinance: they cannot refinance through debt but issuance of new equity only. In reality, equity issuance is seldom inexpensive, as it involves underwriter fees, search costs, or agency costs. Here I abstract away from the mirco-market details by making a reduced form assumption about issuance costs similar to the one used in Bolton et al. (2011): firms must pay a fixed cost ϕ each time they raise new equity; for each unit of net worth raised, there is a marginal cost $1 + \xi$ where $\xi > 0$. Those costs imply that firms will only refinance when their net worth is sufficiently low. In the next section, I will demonstrate how those refinancing costs can be translated into boundary conditions which the solution of the firm's optimization problem must satisfy.

Firms do not engage in either payout or equity issuance when W_t lies in between the payout and the refinancing boundaries. They simply accumulate wealth through investing their entire net worth in land, which is the only durable asset available so far. The dynamics of firm net worth follow

$$dW_t = W_t dR_t = \mu_W dt + \sigma_W dt \ .$$

The drift term μ_W and the diffusion term σ_W will be endogenously determined later. Define $V(W)$ as the value function of a firm given its net worth W. Using Ito's

[4] Firms can also choose to liquidate. I assume liquidation yields value zero, and the expected return from production μ is high enough such that firms always prefer refinancing over liquidation

Lemma, firm value follows the Hamiltonian-Jacobian-Bellman (HJB) equation:

$$\rho V(W) = \mu_W V'(W) + \frac{1}{2}\sigma_W^2 V''(W) \ .$$

An equilibrium consists of the firms' payout and their refinancing decisions that solve the firm's maximization problem subject to the dynamics of firm net worth ((3.1.1)). Moreover, the market for land must clear. Since the total supply of land is fixed to be one, and firms save all their net worth in land, the total value of land exactly equals the aggregate net worth of all firms, that is

$$\int W_t dF(W_t) = P_t \ ,$$

where $F(W_t)$ is the distribution of net worth. This market clearing condition pins down the price of land in equilibrium. This equilibrium is referred to as the *autarky* and serves as the benchmark.

3.1.2 Financial Integration

In autarky, firms can only invest in land, a domestic asset with a fixed supply and risky return. In this section I characterize a new equilibrium after the economy is integrated into the global financial market. Financial integration introduces a new asset: risk free bonds traded on an international bond market. Risk free bonds are supplied by foreign issuers in arbitrary amounts and have a fixed return of r units of rice. Interest rate r is not affected by the demand of domestic firms.

The key assumption in autarky, limited commitment, still applies to domestic firms in the international bond market: they can buy but are restricted from issuing any bond because they are still unable to credibly pledge any future income. In another words, domestic firms are lenders on the international market, consistent with the observed direction of global capital flow from developing countries to developed countries.[5]

Limited commitment has long been recognized as an important friction not only within emerging economies but in the international financial market as well. Kehoe and Perri (2002) argue that international loans are, in general, difficult to enforce due to different legal and cultural environments across countries. Bai and Zhang (2010, 2012) found that despite financial liberalization, countries still vary greatly in their ability to commit to debt payments, and thus there is inefficient global risk sharing as a frictionless model would predict. Here, the implicit assumption is that international bond suppliers have infinite borrowing capacity while domestic firms have zero capacity, which keeps the model simple without losing important qualitative predictions.

The introduction of risk free assets implies a non-trivial portfolio choice problem for firms, who must allocate their net worth between the domestic risky asset, land, and the foreign risk free asset, bonds. Let α_t denote the fraction of firms' net worth invested in land, and $1-\alpha_t$ the holding of risk free bonds. The dynamics of W_t thus follows

$$dW_t = \frac{\alpha_t}{P_t} W_t dR_t + (1-\alpha_t) W_t r dt \; . \tag{3.1}$$

An equilibrium now consists of firms' optimal payout and refinancing decisions plus their portfolio choice α_t, subject to limited commitment $\alpha_t \leqslant 1$, and the dynamics of firm net worth (3.1). The market clearing condition which pins down the price of land is now

$$\int \alpha_t W_t dF(W_t) = P_t \; . \tag{3.2}$$

That is, the total fraction of firm net worth allocated to land equals the total value of

[5] Although the direction of capital flow is the direct result of limited commitment, it can be endogenized using various techniques, for instance with different degrees of risk aversion à la Maggiori (2012). Since the focus of this chapter is not the cause but rather the domestic consequences of capital outflow for emerging economies, the mechanism behind such capital outflow is kept as simple as possible.

land. This equilibrium is referred to as *financial integration* or integration in general.

To summarize the model so far, there is one economy in this model representing emerging markets. Limited commitment implies firms in this economy can only accumulate net worth through one domestic risky asset. Financial integration introduces an international risk free asset with fixed return in which firms can also invest in addition to the domestic risky asset. Firms pay off dividends when their net worth is high and refinance through costly equity issuance when net worth is low.

3.2 Equilibrium Asset Prices and Aggregate Savings Rate

In this section I solve the equilibriums price of land under both autarky and financial integration. I show that, under certain parameter restrictions, land price increases after financial integration. I then define the aggregate savings rate and argue that asset price booms and savings gluts can occur simultaneously after financial integration. I demonstrate that this result comes from the dominating wealth effect of introducing risk free assets to firms' portfolios: less risky portfolios make firms on average wealthier in equilibrium and the total value invested in land is higher. I then discuss the implications of this and present comparative statics.

3.2.1 Asset Prices

To analytically solve the model, I focus on the steady state equilibrium where the price of land P_t does not change over time. The existence of the steady state will be verified later. Dropping the time subscript of P_t implies $dR_t = \frac{dA_t}{P}$, that is the return of land is the return from the output only. Substituting this back to dW_t and using Ito's lemma, the value of the firm solves the HJB equation:

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu}{P} - r \right) \alpha_t W_t + r W_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \, . \tag{3.3}$$

Firms pay out dividends when W exceeds the payout boundary $\overline{W}$ and refinance through equity issuance when W falls below the refinancing boundary $\underline{W}$. These boundaries are determined by the following conditions:

$$V'(\overline{W}) = 1 \,, \tag{3.4}$$

$$V'(\underline{W}) = 1 + \zeta \,, \tag{3.5}$$

and

$$V(\underline{W}) = \phi + (1+\zeta)\underline{W} \,. \tag{3.6}$$

The first condition states that, at the payment boundary $\overline{W}_t$, the shadow value of firm's net worth must equal the marginal value of the dividend, which is one because of risk neutrality. The second condition comes from the fact that refinancing bears a marginal cost of $1+\zeta$ per unit of net worth that firms raise. The third condition is value matching at the refinancing boundary, where a firm's value should have no discontinuity before and after refinancing.

Differentiating equation (3.3) with respect to α_t, the usual first order condition implies

$$\alpha_t = -\frac{\mu P - rP^2}{\sigma^2}\frac{V'(W_t)}{W_t V''(W_t)} \,.$$

This solution of α_t takes the typical Merton form: net return of the risky asset less the risk free interest rate divided by the variance of return and multiplied by the inverse of the coefficient of relative risk aversion.

The HJB equation (3.3) can potentially have multiple solutions. In this chapter I will focus on one particular solution: constant α_t. This means I conjecture and later verify that $V(W_t)$ satisfies $\frac{V'(W_t)}{V''(W_t)} - \frac{1}{\beta}W_t$ for some constant β. Therefore $\alpha = \frac{\mu P - rP^2}{\beta\sigma^2}$ and is independent of time. The HJB equation is then simplified to

$$\rho V = \left(\frac{(\mu - rP)^2}{2\beta^2\sigma^2} + r\right) W_t V'(W_t) \,.$$

This first order ordinary differential equation has the following close-form solution:

$$V(W) = C_1 W^{\gamma} \ ,$$

where C_1 is a constant coefficient to be determined by matching the boundary conditions. The algebraic details can be found in the Appendix. Matching boundary conditions implies the following lemma:

Lemma 3.2.1. $0 < \gamma < 1$ *if* $r < \rho$.

This lemma implies that $V(W)$ is a concave function with slope $1 + \xi$ at the refinancing boundary $\underline{W}$ and 1 at the payout boundary $\overline{W}$. Put differently, even though the firms are risk neutral, they exhibit investment behaviors similar to risk averse investors. Two factors, lack of complete market and costly refinancing, contribute to the risk aversion. Without limited commitment, optimal risk sharing among firms implies complete hedging. Firms have no need for precautionary savings even though investment in land is risky; if firms are allowed to take infinite losses, or if the cost of raising one unit of net worth is exactly one, then the risk of hitting the refinancing boundary does not matter. Firms only consider expected return when choosing between investing in land or risk free bonds because of the absence of aggregate shocks.

The condition $r < \rho$ is necessary for a non-trivial solution: if the market risk free interest rate is higher than firms' discount rate, firms can save through risk free bonds only and postpone the dividend payout forever. Nevertheless, this condition does not necessarily mean that domestic firms are more impatient than international debtors. It can be interpreted as transaction costs from trading on the international market which bring down the effective interest rate of holding foreign risk free bonds for domestic firms.

The first order condition of α_t also implies that in between $\underline{W}$ and $\overline{W}$, the dynamics of W_t follow

$$\frac{dW_t}{W} = \left[\frac{(\mu - rP)^2}{\beta\sigma^2} + r \right] dt + \frac{\mu - rP}{\beta\sigma} dZ_t \ . \tag{3.7}$$

Notice that $\mu > rP$ since $P = \frac{\mu}{\rho}$ is the price of land under the risk neutral measure, which must be strictly higher than the price of land when firms are effectively risk averse.

Equation (3.7) suggests that W_t is a geometric Brownian motion with two reflecting barriers or RGBM for short. This type of processes has closed form stationary distributions which makes solving the market clearing condition of land price possible. Consider a generic RGBM W_t, where $\frac{dW_t}{W_t} = \mu_W dt + \sigma_W dZ_t$ between an upper reflecting barrier $\overline{W}$ and a lower reflecting barrier $\overline{W}$. Let $\eta \equiv \frac{2\mu_W}{\sigma_W}$, Zhang and Du (2010) show that the stationary distribution of W_t is given by the density function:

$$f(W) = \frac{\eta - 1}{\overline{W}^{\eta-1} - \underline{W}^{\eta-1}} W^{\eta-2} \ . \tag{3.8}$$

This density function is a power function that can be easily integrated to obtain its expectation:

$$E(W) = \int_{\underline{W}}^{\overline{W}} \frac{\eta - 1}{\overline{W}^{\eta-1} - \underline{W}^{\eta-1}} W^{\eta-1} dW = \frac{(\eta - 1)}{\eta \left(\overline{W}^{\eta-1} - \widehat{W}^{\eta-1} \right)} \left(\overline{W}^{\eta} - \underline{W}^{\eta} \right) \ . \tag{3.9}$$

Since W_t is an RGBM under both autarky and financial integration, $E(W)$ can be calculated in the same way for different endogenous drift and diffusion terms under both equilibriums. This allows me to analytically solve for and compare the equilibrium land price P. Plugging $E(W)$ into the market clearing condition of autarky, $E(W) = P$, and the market clearing condition of financial integration,

$\alpha E(W) = P$, and noticing that

$$\alpha = \frac{\mu P - rP^2}{\beta\sigma^2}$$

is a function of P in the steady state yield two systems of equations involving P and β, where β is an implicit function of P. More details, including the explicit form of the systems of equations, are given in the Appendix.

Deriving the equations for P allows me to examine the circumstances under which asset price booms and savings gluts can arise simultaneously. An asset price boom is defined as a higher domestic asset price P after financial integration than under autarky. The following proposition gives the condition for its occurrence:

Proposition 3.2.1. *(Asset Price Boom) Let P^A be the price of the domestic risky asset in the stationary equilibrium under autarky and P^I be the price under financial integration. Defining*

$$\frac{3r + (2r-1)\rho}{\rho - r} \leqslant \frac{2\phi\,(1+\xi)^2 - 1}{(\xi+2)(\xi+1)}$$

as Condition 1, then when Condition 1 holds,

$$P^I > P^A\ .$$

Proposition 3.2.1 is the result of two conflicting effects on the demand of the domestic risky asset brought about by the access to risk free bonds. On one hand, there is a *substitution effect*, where risk free bonds reduce the demand of land, as they can both serve as a means of precautionary savings for firms to avoid costly refinancing. This is reflected by $\alpha < 1$ in the financial integration equilibrium. On the other hand, there is also a *wealth effect*, where adding risk free bonds to firms' portfolios reduces the overall riskiness of firms' investments. Less likely to hit the refinancing boundary, firms are in general richer in the steady state. Consequently, when the wealth effect dominates, even though firms only invest partially in land, the total value of net worth saved through land is higher and so is the price of land.

The wealth effect can be further clarified through the stationary distribution of RGBM in (3.8), which implies the cumulative distribution density (CDF) function for any $\widehat{W} \in (\underline{W}, \overline{W})$ in the steady state is

$$F^s(\widehat{W}) = \frac{\widehat{W}^{\eta-1} - \underline{W}^{\eta-1}}{\overline{W}^{\eta-1} - \underline{W}^{\eta-1}} , \tag{3.10}$$

where $s \in A, I$ denotes the state: autarky or financial integration. This density function is increasing in η and thus decreasing in σ_w, holding all other variables constant. In the autarky equilibrium, σ_w equals $\frac{\sigma}{P}$, while in the integration equilibrium it equals $\frac{\alpha\sigma}{P}$. That is, smaller share of net worth invested in land implies less likelihood of W_t dropping. As a result, the stationary distribution of W_t is skewed more to the left, and more firms stay near the pay-off boundary in equilibrium. If the distributional change is sufficiently strong, the aggregate wealth of the economy after financial integration is much larger than under autarky, and the fraction of wealth invested in land is also higher.

Proposition 3.2.1. is derived under the parameter restriction in Condition 1, which in all its complexity can be effectively read as refinancing cost must be large enough for the wealth effect to dominate. Economically, higher refinancing costs imply that firms are more concerned about hitting the refinancing boundary, thereby inducing stronger risk aversion which leads to a larger effect from risk free precautionary savings. Mathematically, higher refinancing costs imply a wider space between the refinancing and payout boundaries for firms to accumulate net worth, allowing more room for the distribution of firm net worth to change during financial integration.

Although the extent to which Proposition 3.2.1 applies in reality depends on the actual level of financing costs, there are two reasons why the condition of sufficiently high financing costs is likely to be met in practice. First, there is substantial empirical evidence that external financing is indeed costly in emerging economies.La Porta et al. (1997, 1998, 2000) show that countries with less formal protection of creditor rights and a poorer legal system of commitment have smaller domestic loans, narrower debt markets, and greater difficulty in raising external financing. Allen et al. (2005), Ayyagari et al. (2010) find costly informal financing in a country with less developed formal financial markets. Secondly, one should take into account that Condition 1, derived as a sufficient condition, can be potentially over restrictive. While the necessary conditions are hard to derive explicitly due to algebraic complexity, the latter part of this section offers some numerical examples of comparative statics that serve as reference for the true range of the parameter space. Overall speaking, sufficiently high refinancing costs are required for the conclusion in Proposition 3.2.1 to hold, but such requirement is found numerically plausible and is largely consistent with what we observe for most emerging economies.

3.2.2 Aggregate Savings Rate

In addition to equilibrium asset price, the closed form distribution of firm wealth allows me to calculate the aggregate savings rate and derive the conditions under which a savings glut can occur. Aggregate savings rate is defined as the proportion of wealth not consumed out of the total social wealth which equals the sum of two parts: the total value of assets, $E(W)$, and the total value of consumption produced by the assets. In autarky, the only asset is land, and total consumption equals μ given no aggregate shock. The aggregate savings rate

$$S^A \equiv \frac{E(W)}{\mu + E(W)} \ . \tag{3.11}$$

After financial integration, total consumption equals $\mu + r(1-\alpha)E(W)$, where $r(1-\alpha)E(W)$ corresponds to the amount of interest earned from risk free bonds. The aggregate saving rate

$$S^I \equiv \frac{E(W)}{\mu + (1 + r(1-\alpha))E(W)} \ . \tag{3.12}$$

Both (3.11) and (3.12) can be computed given the value of $E(W)$ and α from the previous section. A savings glut is defined as $S^I > S^A$, that is, savings rate after financial integration exceeds that under autarky. The following proposition gives a sufficient condition for a savings glut:

Proposition 3.2.2. *(Savings Glut) Defining*

$$(1+\zeta)r < 4\mu$$

as Condition 2, then under Conditions 1 and 2

$$S^I > S^A \ .$$

The intuition behind Proposition 3.2.2 follows closely from the intuition of Proposition 3.2.1 in that the savings glut also arises from the dominating wealth effect of introducing risk free assets into firms' portfolios. Assume, for a moment, that r is small, such that total consumption before and after financial integration stays roughly the same. From its definition, it is clear that the savings rate is higher if the total value of assets $E(W)$ is higher. Given the satisfaction of Condition 1, Proposition 3.2.1 implies that $P^I > P^A$, which further implies that $E^I(W) > E^A(W)$, since land value represents the total value of assets in autarky but only a fraction of that under financial integration. The increase in total asset value results directly from the more left-skewed steady state distribution of firm net worth, which is the wealth effect discussed following Proposition 3.2.1.

Condition 2 is necessary for Proposition 3.2.2 in addition to Condition 1 to ensure that the increase in asset value exceeds the increase in consumption, which takes the form of interest rate payments for the risk free bonds. Intuitively, in order for the additional consumption to be small, the return from risk free bonds, r, cannot be too high compared to the return from land, μ. Condition 2 gives the quantitative relationship that the two parameters need to satisfy. Furthermore, the return from the two assets play different roles in determining land price and the savings rate. For land price, the smaller r relative to μ, the weaker substitution effect between land and risk free bonds, as the demand for risk free bonds is lower. For the savings rate, smaller r implies a more dominating effect of the change in net worth distribution and total asset value relative to the change in total consumption, which is necessary for a savings glut to emerge.

Condition 2 and Condition 1 can be satisfied simultaneously. Fixing μ, r and ζ such that Condition 2 is met, there is always some ϕ large enough such that Condition 1 is also met. This again highlights the important role of substantial refinancing costs in this model. However, it is equally important to keep in mind that for analytical tractability, Proposition 3.2.2 is derived under the validity of Proposition 3.2.1—potentially a very limited area in the entire space of the necessary and sufficient conditions. Comparative statics in the latter part of this section shows a much broader space of parameters in which Proposition 3.2.1 and 3.2.2 both hold.

The demand for foreign risk free assets in this model is akin to that in other studies such as Mendoza et al. (2009) and Maggiori (2012), which also feature endogenous change in savings rate after financial integration. In Mendoza et al. (2009), an emerging country and a developed country differ in the completeness of their domestic markets: the emerging country is not able to trade contingent claims because of its agents' ability to falsify their true income, whereas the developed country has complete credit markets. Although similar to the assumption made in this chapter, the key force that determines the change in the savings rate of the emerging country following the integration of the two economies is the difference in interest rates between the two economies in autarky. The emerging country has a lower interest rate and consequently higher asset prices in autarky because its investors have a stronger motive for precautionary savings. In Maggiori (2012), global financial imbalance is triggered by the asymmetry in risk aversion between the developed country (US) and the emerging economy (rest of the world, ROW). Intermediaries in the ROW face a non-negative equity constraint while the US intermediaries do not, making the latter better at absorbing investment risk. As a result the US holds most of the risky assets in the world and the ROW mostly risk free debt issued by the US intermediaries. In both of these studies and in this paper, agents of the emerging economy have strong precautionary savings motive since they cannot completely hedge investment risk due to their domestic market frictions. This precautionary savings motive increases the aggregate savings rate of the emerging economy once foreign supplies of risk free assets become available.

This paper differs from existing studies in the mechanism that jointly determines both asset prices and savings rate. Aside from the precautionary savings motive raising the demand for risk free assets, existing studies usually also predict less demand for domestic risky assets from the investors of emerging economies who substitute risky assets for risk free assets. In contrast, this chapter argues that in addition to the substitution effect of higher precautionary savings through risk free assets, there is also a wealth effect that can dominate as long as domestic firms are sufficiently risk averse. A dominating wealth effect pushes up both the price of the domestic risky asset as well as the overall value of firm wealth.

Conclusions in this chapter contribute to the debate over whether foreign direct investment (FDI) is the major factor responsible for the asset price booms in some emerging economies during the 1990s. The existing literature predicts that as domestic investors substitute domestic risky assets with foreign risk free assets, the domestic risky assets are usually being held in the form of FDI by foreign investors, who have the comparative advantage of making risky investments due to a more developed financial market. The price of domestic risky assets increases only if the demand through FDI exceeds the supply due to the substitution effect. However, despite the FDI's rapid growth in many emerging economies, it is still considerably smaller than the domestic investment. This chapter casts doubt on FDI as the major cause for domestic asset price increase by showing that asset price booms can emerge solely as a result of increased domestic demand. This sounds a caveat to policy makers who aim at regulating FDI for the purpose of decelerating asset price booms.

This model also generates an endogenous *liquidity glut*, a concept proposed by Rajan (2006b), who argues that excessive liquidity and shortage of physical asset in many emerging economies contribute significantly to their asset price booms. Advocators of the liquidity glut theory usually attribute excessive liquidity to imbalanced growth between production and investment. However, such argument is controversial given the rapid industrialization in several emerging economies. [6] In contrast, this chapter predicts that more wealth can be invested in the domestic physical asset, land, when the wealth effect of financial integration dominates. Excessive liquidity can rise endogenously due to precautionary savings and reduced overall investment risk.

The stationary distribution of firms' wealth in this model, given by equation (3.10), takes the form of a power function. This provides an potential explanation for the cross-sectional distribution of firm size, which likewise closely resembles a power function. Ai et al. (2012) calculate that the number of Compustat firms

[6] For example, the investment rate in China post the 1992 economic reform is maintained around 39 percent accordingly to Song et al. (2012)

with more than N employees is roughly $N^{-1.06}$. They establish a theory featuring such power distribution of firm size by explicitly modeling agency frictions. In the model of this chapter, W_t which measures the value of firms' total assets can be naturally interpreted as firm size. W_t follows a power distribution in the steady state equilibrium in this model because of the stationary distribution of a reflected geometric Brownian motion. Although built intentionally for emerging economies, the model in this chapter can potentially be calibrated to fit the US data as well. Note that while imperfect market is crucial for the asset price and savings rate movements to be consistent with observation, such assumption is not essential for a power function form of stationary distribution, which only requires a geometric Brownian motion with two reflecting barriers.

3.2.3 Comparative Statics

How do land price and savings rate react to exogenous changes in productivity, international interest rate, or financing costs? In this section I study the comparative statics of several exogenous variables through numerical simulations.[7] The variation in exogenous variables can be interpreted as unexpected aggregate shocks or sudden regulatory changes. Comparative statics serve as a first step for further investigation of those regulatory changes and provide a reference for the parameter space for the main propositions above.

I consider four parameters that generate the most interesting implications from this model: the expected return of land, μ; the variance of the return on land, σ; the interest rate r of risk free bonds; and the marginal cost of raising new equity, ξ. For each parameter, I illustrate equilibrium land prices P^I and P^A and their differences, savings rates S^I and S^A and their differences, firms' effective degree of risk aversion

[7] Although analytical sufficient conditions can be derived for some parameters, they are in general overly restrictive and difficult to interpret.

β^I and β^A and their differences, and portfolio choice α.

Return of the Domestic Asset

I first consider comparative statics for μ, the expected return from investing in land. Figure 4 shows the equilibrium prices of land, P^I and P^A, as functions of μ. Both P^A and P^I are increasing functions in μ, which is very intuitive: higher expected return makes land more valuable in both autarky and financial integration. The difference between P^I and P^A is always positive, indicating the robustness of Proposition 3.2.1, which is analytically derived under sufficient conditions only.

Figure 4 also shows the difference between P^I and P^A diminishing as μ becomes larger, an interesting observation illustrating the joint effect of the two major forces driving land price in this model: the substitution effect and the wealth effect of introducing risk free bonds. A more intuitive reading of Figure 4 in understanding this joint effect is that the difference between P^I and P^A expands as μ becomes smaller. On the one hand, the substitution effect is stronger for smaller μ. This is shown in Panel D of Figure 4, where the portfolio weight of land, α, is an increasing function in μ. On the other hand, return on land changes firms' degrees of risk aversion β^A and β^I, leading to different valuation of land in equilibrium. As Panel C demonstrates, β^A rises significantly as μ becomes smaller, while β^I stays relatively flat. This shows that firms are much more risk averse when μ is low, which decreases their valuation of land. This is intuitive in that a lower μ makes it easier to reach the refinancing boundary. Altogether, the difference in land price before and after financial integration suggests a dominating wealth effect when μ is small, the conclusion drawn in Proposition 3.2.1.

Figure 4 also shows the pattern in the savings rate. Consistent with Proposition 3.2.2, Panel B shows that the savings rate is always higher after financial integration. However, the bigger the μ, the smaller the difference, as S^A is an increasing function in μ. This is again evidence of a dominating wealth effect, which increases the total asset value, $E(W)$, in the savings rate equation (3.11). Just as in the case of land price, μ has little effect on the post financial integration savings rate. This is because return on the risky asset matters less to firms when risk free precautionary savings are available.

Risk of the Domestic Asset

The second parameter explored here is σ: the return volatility of land which measures the risk of domestic investment. The value of this parameter is closely related to financial stability and business cycles, which can fluctuate quite dramatically in many emerging economies. Figure 5 illustrates the effect on land price and savings rate of changing σ. Both P^I and P^A are lower when σ is larger, as investing in a more risky asset requires a higher premium. At the same time, S^I and S^A move in opposite directions. As a higher risk of land in autarky naturally implies lower savings through land, the increasing risk also reinforces the precautionary savings motive, increasing savings under financial integration when the risk free asset becomes available.

Panel D in Figure 5 shows that α drops significantly when σ increases. The decline in α can be viewed as an *"investment slump"*, that is, a protracted decline in physical investments observed for many of emerging economies that experience savings gluts. Existing studies on this phenomenon, for example Kramer (2006) and IMF (2007), attribute to it the observed direction of global capital flows. However, these studies normally exclude countries without an open market, such as China, and usually also predict a slump in the domestic asset prices. The result in this section not only shows that a slump in domestic investment can follow from increased investment risk but also that it can trigger asset price booms if a country simultaneously opens up its financial market.

Risk Free Interest Rate

Next I consider the risk free interest rate r. Recall that risk free bonds are supplied in the international market by foreign issuers. Domestic firms are price takers and their investments have no effect on the return of those bonds. However, the international bond market is by no means stable. It is constantly being affected by the national credit status and unexpected extreme events. Caballero et al. (2008) find that the global risk free interest rate has been declining since the 1980s. However, they focus only the impact on the US market. The comparative statics in this part provide, for the first time, theoretical predictions on the dynamics of asset price and savings rate for emerging markets as they experience global interest rate changes.

Figure 6 plots land price and savings rate as functions of r. P^A, S^A, and β^A are constant in r which has no effect on the variables in autarky. Meanwhile, P^I increases in r, a result of two effects working against each other: on the one hand, the substitution effect is stronger when r is bigger, which is reflected in the declining α in Panel D. On the other hand, higher interest rates improve firms' returns from precautionary savings, lowering their degree of risk aversion. This is illustrated in Panel C, where β^I declines in r. A similar argument applies to the impact of r on the savings rate, as shown in Panel B of Figure 6. Savings are increasing in r after financial integration. According to the savings equation (3.12), higher r implies higher aggregate consumption but also higher aggregate value of assets held by firms due to a stronger wealth effect. Savings increasing in r again suggests a dominating wealth effect.

Refinancing Cost

Last but not least, I demonstrate in Figure 7 the effect of marginal refinancing cost, ξ, which is the key variable determining the magnitude of wealth effect in this model. Not surprisingly, P^I and P^A are decreasing in ξ, because higher refinancing costs make firms more risk averse and lower their valuation of the risky asset. The difference between P^I and P^A is increasing in ξ because P^A drops faster than P^I as refinancing cost increases. This result can be predicted by Condition 1 and 2, the sufficient conditions under which Proposition 3.2.1 and 3.2.2 are derived, as higher refinancing costs make those conditions easier to satisfy.

In this section's discussion of comparative statics, parameter values given in Figure 2 are used as the benchmark. These numbers are largely consistent with standard investment models, such as Bolton et al. (2013), but subject to several minor modifications: I choose higher values for σ and ζ to reflect the fact that in most emerging economies, domestic investment is highly risky and external financing rather difficult.

3.3 Welfare Analysis

The massive savings through foreign risk free assets in emerging economies have triggered much political and economical debate over whether they impede domestic investment and growth. Such debate calls for the study of the welfare implications of firms' portfolio choice after financial integration. In the decentralized equilibrium, firms make portfolio choices between risk free bonds and land, taking land price as given. Firms do not internalize the impact on land value that their portfolio choice can have through the aggregate distribution of firm wealth. In a market with imperfect commitment where market allocation of firm wealth is constrained, inefficient portfolio choices and suboptimal equilibrium asset prices can arise.

The welfare analysis in this section reveals that decentralized portfolio choices are indeed inefficient: the socially optimal portfolio places less weight on the domestic risky asset, land, and more weight on foreign risk free bonds. The socially optimal equilibrium also displays lower asset price and savings rate under financial integration, suggesting that the asset price boom and savings glut in Section 4 are too high.

The socially optimal portfolio choice can be achieved through a tax on land.

3.3.1 The Social Planner's Problem

I conduct the welfare analysis in a similar manner to Lorenzoni (2008): a social planner maximing the same objective function as firms must face the same limited commitment constraint and refinancing cost that firms do. In other words, the social planner has no power to make either intratemporal or intertemporal resources reallocations among firms. The only additional power the social planner has is knowing the relation between portfolio choice and equilibrium asset price, which the social planner takes into consideration when making his portfolio choice decisions.

The social planner's value function $V(W_t)$ solves the following HJB equation:

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu}{P} - r \right) \alpha_t W_t + r W_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \ ,$$

subject to the dynamics of W_t given by (3.1) and boundary conditions (3.4), (3.5), and (3.6). Most importantly, the social planner takes into account how equilibrium land price P is determined. That is, there is one additional constraint to the social planner's problem: the market clearing condition,

$$\int \alpha_t W_t dF(W) = P$$

The same technique used to solve the firm's problem in the previous section can still be applied here. I leave the details of the solution to the appendix and move directly to the following proposition that summarizes the finding:

Proposition 3.3.1. *(Constrained Efficiency) Let α^* be the solution to the social planner's problem. Then, under Condition 1,*

$$\alpha^* < \alpha^I \ .$$

Proposition 3.3.1 states that the social planner's optimal portfolio places a smaller weight on land and a larger weight on risk free bonds relative to firms' decentralized decisions. The intuition can be best understood through a reverse argument in the proof of Proposition 3.2.1 which is detailed in the Appendix. First of all, the degree of firms' risk aversion is decreasing in the proportion of wealth allocated to the risky asset, since risk aversion comes from the concern of hitting the refinancing boundary and having to costly raise new capital, a portfolio with less weight on land reduces this probability of refinancing. Mathematically, $\frac{\partial \beta}{\partial \alpha} < 0$. Secondly, less risk averse firms value the risky asset more, that is $\frac{\partial P}{\partial \beta} > 0$. Altogether, by substituting some investment in land with investment in risk free bonds, firms place less weight on land after financial integration. This substitution implies a higher land price and lower marginal return from land, a pecuniary externality which firms do not internalize when making decentralized decisions. However, when this externality is taken into consideration, the social planner is aware of the lower return from land and consequently allocates more value to the risk free bonds.

The pecuniary externality and inefficient portfolio choice come from a combination of incomplete market due to limited commitment and anonymous spot trading of assets. The welfare impact of such a market structure in this chapter resembles the studies of Kehoe and Levine (1993), Lorenzoni (2008) etc, who find that market frictions in general lead to inefficient asset allocation and firm behaviors, potentially causing macroeconomic cycles. In this model, inefficient portfolio choices lead to suboptimal asset prices and distribution of firm wealth. Eisfeldt and Rampini (2006) show that redistribution of firm wealth is strongly pro-cyclical, whereas the socially optimal reallocation should be counter-cyclical. The model used in this chapter is flexible and can be easily modified to incorporate aggregate shocks and to study the implications for business cycles.

Proposition 3.3.1 sheds light on the ongoing political debate over the appropriate level of domestic asset prices in some emerging economies. The increase in asset prices after financial integration improves overall efficiency compared to autarky, as the distribution of firm net worth is more skewed to the left and firms are on average wealthier. However, the *level* of asset price increase does not achieve efficiency. The socially optimal portfolio derived in Proposition 3.3.1 invest less proportion in land. Meanwhile, this also implies a less risky portfolio and hence an even more left-skewed stationary distribution of firm wealth as well as higher aggregate value of investment in land. The overall effect on land price depends on the relative strength of these two opposing forces. The socially optimal land price P^* appears lower than P^I, because the effect of less weight on land dominates. This suggests that even though higher asset prices under financial integration reflect a welfare improvement over autarky, the degree of such asset price booms is inefficiently high.

The fact that the social planner saves more through foreign risk free assets provides an explanation to the large holdings of US treasury securities by the government of many emerging economies, such as China. Such government investment strategy is highly controversial. Critics disparage the large portion of national wealth, mostly procured through the profits of state-owned enterprises, being saved in the form of foreign debt rather than invested in domestic assets. The analysis in this section shows that should the government transfer its investment from foreign risk free assets to domestic risky asset, the price of of domestic assets, such as land, would be even higher and more inefficient. Acting like a social planner, the government allocates more resources to holding low-risk sovereign debt because a decentralized economy would otherwise engage in insufficient precautionary savings.

The use of *constrained efficiency* refers to the social planner's limited ability to make intratemporal or intertemporal reallocation of resources. It was proposed as an useful tool to study the minimal set of conditions under which competitive allocation can be improved. It also helps the analysis of various regulatory policies imposed under market constraints. Moreover, under the constrained efficiency framework, interventions aimed at one market (e.g., the asset market) can have important interactions with the equilibrium outcome of other markets. One application of the welfare analysis, the planner directly intervening in the asset market through taxes and subsidies, is of particular interest to policymakers and is detailed in the next section.

3.3.2 Implementing the Social Planner's Choice

An inefficient decentralized equilibrium naturally leads to the discussion of possible government policies to restore asset prices and wealth distribution to efficiency. The most common policy that implements the socially optimal choice in this type of constrained efficiency model is a tax or subsidy that internalizes the pecuniary effect of individual firms' portfolio choices on equilibrium asset prices. Consider a tax τ which the government imposes on holding land.[8] Let $\alpha^I(\tau)$ be firms' equilibrium portfolio choice given tax rate τ, the firms' value function now solves

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu}{P} - r - \tau \right) \alpha_t W_t + r W_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \,. \quad (3.13)$$

Applying the same technique in Section 3.3 and comparing the solution to the constrained efficient portfolio choice α^* in Proposition 3.3.1, I obtain the following proposition:

Proposition 3.3.2. *(Optimal Tax Rate) There exists $\tau > 0$ such that, under Condition 1, $\alpha^I(\tau) = \alpha^*$.*

Proposition 3.3.2 is quite intuitive given that the socially optimal portfolio places less weight on land. As investment in risk free bonds reduces the riskiness of firms' portfolios, equilibrium price of land increases as a result of a dominating wealth

[8] A negative τ can be interpreted as a subsidy.

effect, which implies a lower return on land. Firms in the decentralized economy with anonymous spot trading of assets do not take into account such externality and hence miscalculate the actual return of land. A tax on land suppresses its demand and brings the decentralized decision to the socially optimal choice. This tax can be also be inferred from comparative statics in Section 4.3, where the portfolio weight on land α is an increasing function in the return of land, μ.

Proposition 3.3.2 offers a recommendation to policymakers of emerging economies who want to regulate domestic asset prices and savings rate. Public supply can alleviate the problems of competitive equilibrium allocation when the competitive market is subject to frictions. For emerging economies, the welfare analysis in this section suggests potentially over-investment in domestic risky assets and under-investment in precautionary savings. A tax on the risky investment provides a less risky investment environment and alleviates excessive savings and price booms, improving overall welfare.

3.4 Heterogenous Firms and Total Factor Productivity

This chapter is motivated by the global financial account imbalance resulting from financial integration. Capital flowing from emerging to developed countries represents asset reallocation via the international market. While cross-country asset reallocation due to financial integration has been discussed extensively by existing studies, the domestic asset reallocation is less well understood.

This section fills that void by introducing heterogeneity into firms' productivity, and reveals how asset reallocation generated by heterogeneous productivity can amplify asset price booms and savings gluts after financial integration. It also demonstrates how financial integration improves the aggregate productivity of the economy, but this improvement has a non-monotonic relationship with the degree of heterogeneity. This produces several testable hypotheses explaining cross-country differences in asset price and savings rate.

Research has argued heterogeneous productivity and asset allocation to be potentially important determinants of many economic indices, especially among emerging economies. Gourinchas and Jeanne (2013) find that asset allocation is related to country specific savings rates. Song et al. (2012) establish a two-sector economy model, where asset allocation contributes to the level of growth, investment as well as current account balance. This section extends these studies to financial integration-induced asset reallocation and demonstrates how it impacts domestic asset prices, savings rate, and total factor productivity.

3.4.1 Heterogenous Productivity and Asset Prices

I alter the model to incorporate heterogeneous productivity in the simplest way. Let $dA_t = \mu_s dt + \sigma dZ_t$, where $s \in (l, h)$ is a binary state variable denoting firms' productivity: either low type μ_l or high type μ_h. The distribution of high productivity firm has measure π. This simply means that the dynamics of firm wealth is

$$dW_t = \frac{\alpha_t}{P_t} W_t dR_{s,t} + (1 - \alpha_t) W_t r dt \ . \tag{3.14}$$

In the steady state, firm's value function follows

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu_s}{P} - r \right) \alpha_t W_t + r W_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \ ,$$

subject to the dynamics of W_t for both types of firms and the same boundary conditions (3.4), (3.5) and (3.6).

At the same time, the market clearing condition becomes

$$\pi \int \alpha_h W_h dF_h(W) + (1 - \pi) \int \alpha_l W_l dF_l(W) = P \ ,$$

which stems from an aggregate savings in land by both types of firms that equaled the total value of land.

In this binary state model, the degree of heterogeneity across firms can be measured by the range of μ_s. I compute numerical examples of mean-preserving spreads of μ by setting $\pi = 0.5$ and increasing μ_h, while decreasing μ_l to keep the same average productivity as in Figure 3. The higher the μ_h is, the larger the dispersion of productivity and degree of firm heterogeneity.

Figure 8 shows land prices as a function of productivity heterogeneity. Both P^A and P^I decrease as the dispersion of productivity expands. Notice that P^I is still higher than P^A, consistent with the prediction in Proposition 3.2.1. However, their difference increases with higher degree of heterogeneity. That is, heterogeneous productivity amplifies the asset price booms resulting from financial integration. Recall that in Section 5, such booms are inefficiently high, in other words heterogenous productivity exacerbates the inefficiency.

Compared to the baseline homogenous productivity model, there are two additional channels through which land price varies with firm heterogeneity: first, whereas homogenous firms hold the same proportion of risky assets in their portfolios, heterogenous productivity induces asset reallocation between the two types of firms with different returns from land. Whether higher returns imply more investment in land depends on firms' intertemporal elasticity of substitution which, in this model, is induced by costly refinancing. Panel B of Figure 8 plots α_s, the portfolio weights on land, for firm type. High productivity firms invest a lower proportion in land than low productivity firms. Their investment in land is also a decreasing function of productivity, implying an elastic intertemporal substitution effect. Higher return from land increases firms' willingness to trade future dividend payout for precautionary savings that lowers investment risks, which also explains why less productive firms invest more in land.

The second channel through which heterogeneity affects land prices is change in firms' effective degree of risk aversion, which determines the valuation of any risky asset given its return. Figure 8 shows that under both autarky and financial integration, firms' risk aversion coefficient β increases for low productivity firms but decreases for high productivity firms. In the steady state equilibrium, not only is there a transfer of shares of land captured by α_s across firms with different productivity levels but also a transfer of firm wealth from low to high productivity firms. While both types of firms are subject to wealth effect after financial integration, it is stronger for lower productivity firms as they are more risk averse.

To sum up, introducing firm heterogeneity amplifies the boom in land price after financial integration through the transfer of firm wealth between low and high productivity firms and asset reallocation, with the latter dominating. The incremental investment in land by low productivity firms cannot offset the reduction in investment by high productivity firms, lowering the aggregate value invested in land.

Heterogeneous productivity also amplifies the savings glut, as shown by Figure 9. The breakdown of savings for low and high productivity firms is also shown. The trends in savings follow closely those of β_s: high productivity firms save less than low productivity firms, and their savings also decrease in productivity. The same intuition for elastic intertemporal substitution applies here as well. Savings decrease faster in productivity under autarky than under financial integration, because firms are more risk averse in the former equilibrium. As heterogeneity grows, high productivity firms in autarky reduce savings the most, causing a wider gap in aggregate savings before and after financial integration.

Results in this section regarding heterogeneous productivity provide the following empirical predictions: cross-country differences in asset prices and savings rate can be explained by heterogeneity in countries' domestic firm productivity, and even more so by the productivity level of the least productive firms. Research has documented severe market frictions in many emerging economies that hinder technology diffusion and create heterogeneity among domestic firms. The same market frictions can also magnify the growth in asset prices and savings rate once when these economies are integrated into the global financial market.

3.4.2 Financial Integration and TFP Changes

The effect of heterogeneity on asset prices and the savings rate described in the previous section stems from asset reallocation in an imperfect market due to financial frictions, namely, limited commitment and costly financing. With firms now varying in the level of productivity, the model is equipped to study the implication of heterogeneity for aggregate productivity of the economy and answer whether financial frictions cause asset misallocation, a question heavily debated in literature. On the one hand, Buera et al. (2011), Hsieh and Klenow (2009), Moll (2013) argue that asset misallocation plays a significant role in explaining cross-sectional productivity discrepancies among different industries and countries. On the other hand, empirical studies such as Midrigan and Xu (2011) argue that when matched to micro-level data, financing friction cannot generate much asset reallocation and hence neither large cross country differences. In short, there is no consensus on how and to what extent financial frictions matter for productivity.

I address this question by asking whether financial integration, which improves the completeness of financial markets by introducing a risk free asset, also improves the aggregate productivity of the economy. It is straightforward to define total factor productivity (TFP) for the domestic asset land in this model: since firms with expected productivity of μ_s invest α_s in land and there is no aggregate shock, TFP equals total output divided by the quantity of land which is fixed at one unit:

$$TFP = \sum_{s=l,h} \pi_s \int \mu_s \frac{\alpha_s}{P_t} W_s dF_s(W) \ . \tag{3.15}$$

In the steady state equilibrium, α_s and P_t are constant. In autarky, $\alpha_s = 1$ for all firms regardless of their productivity. TFP is simply the average of μ_s weighted by the relative distribution of firm wealth.

Figure 10 demonstrates numerical examples of TFP changes following financial integration as a function of mean preserving productivity spread. Three major observations can be drawn: first, TFP in both autarky and under financial integration is decreasing in heterogeneity. Secondly, compared to autarky, TFP is always higher after financial integration. Thirdly, and most interestingly, the increase in TFP is non-monotonic in heterogeneity.

The observations drawn from Figure 10 can be explained by two effects of financial integration: reallocation of land, and redistribution of firm wealth. Recall in the previous section, low productivity firms hold more risky assets than the high productivity firms because of elastic intertemporal substitution. Since more land is reallocated to low productivity firms, TFP is lower with more heterogeneity. Meanwhile the increase of TFP after financial integration suggests the total wealth high productivity firms invest in land more than offsets the lower proportion of land held in their portfolio. The reallocation of land and redistribution of firm wealth work in opposite directions. Mathematically, both α_s and W_s are endogenously determined in equation (3.15). α_h decreases while W_h increases in μ_h. The combination produces non-monotonic TFP changes as a function of firm heterogeneity.

This finding sheds light on the mixed empirical evidence for the relationship between financing frictions and cross-sectional TFP dispersion. Previous research such as Midrigan and Xu (2011) and Moll (2013) argue that the persistence of productivity shocks is key to understanding such relationship. More persistent productivity shocks imply that financial frictions play a less critical role, as firms can better anticipate their future financial need and save accordingly to grow out of their financial constraint. In this model, productivity is perfectly persistent without any aggregate uncertainty, whereas the change of TFP is still non-monotonic in the dispersion of firm productivity. This implies a more intricate relationship than what has been proposed so far and warrants future analyses.

3.5 Remarks

Financial integration is important for understanding many market behaviors of e-merging economies. This chapter provides a model that features endogenous asset prices and savings rate, and argues that the simultaneous observation of savings g-luts and asset price booms in emerging economies can be the result of a dominating wealth effect brought by global financial integration. The critical assumptions are limited commitment implying incomplete market and costly refinancing which in-duces firms to balance between risky investment and risk free precautionary savings. Welfare analysis reveals that firms' precautionary savings is inefficiently low, leading to overly inflated asset prices that can also be exacerbated by heterogeneity in firm productivity.

The model of this chapter can be extended in several ways, the most important and interesting of which is perhaps calibration using data from emerging economies. A calibrated model can demonstrate how much of the increase in asset prices and savings is attributable to the wealth effect highlighted in this model. It is also interesting and important for obvious policy reasons to calculate the optimal level of tax be impose in order to implement the socially optimal portfolio choice.

The assumption of fixed asset supply is widely used in micro-founded asset pricing models. Thought it allows analytical tractability, the model necessarily becomes silent on capital reproduction. However, the dynamics of capital reproduction can shed light on the credit cycles of emerging economies during their development, which would certainly be interesting for future study.

Appendix A

Proofs

A.1 Proofs for Chapter 2

Proof of Proposition 2.2.1: This optimality condition for a single state environment is identical to the baseline model in DeMarzo and Sannikov (2006) if the commitment constraint $V''(\overline{W}) \geqslant L$ is not binding, so I will only focus on the case when such condition is violated. I use superscript L to denote variables and functions for the limited commitment case.

Define the social benefit function as $F(W) = W + V(W)$, which satisfies

$$F''(\overline{W}) = \frac{rF(\overline{W}) + (\gamma - r)\overline{W} - \mu}{\frac{1}{2}\lambda^2\sigma_s^2} \, .$$

When the principal's participation constraint is binding, $F^L(\overline{W}) = L + \overline{W}^L$ implying

$$F^{L\prime\prime}(\overline{W}^L) = \frac{rL + \gamma\overline{W}^L - \mu}{\frac{1}{2}\lambda^2\sigma_s^2} \, .$$

Suppose $F^{L\prime\prime}(\overline{W}^L) > 0$, that is, $rL + \gamma\overline{W}^L > \mu$, this implies that $V^{L\prime\prime}(\overline{W}^L) > 0$. Since $V^L(\overline{W}^L) = L$, $rV^L(\overline{W}) + \gamma\overline{W}^L > \mu$. Compare this result to the case of full commitment, where $rV^(\overline{W}) + \gamma\overline{W} = \mu$. If $\overline{W}^L < \overline{W}$, since $rV(W) + \gamma W < \mu$ for all $W < \overline{W}$, it must be that $rV(\overline{W}^L) + \gamma\overline{W}^L < \mu$, which implies $V(\overline{W}^L) < V^L(\overline{W}^L)$. However this is a contradiction since $V(W) \geqslant V^L(W)$ for every W. If, on the other hand, $\overline{W}^L > \overline{W}$, but $V^L(\overline{W}^L) = L$ and $V(\overline{W}) < L$, which implies that $V^L(\overline{W}^L) > V(\overline{W}^L)$ again, contradiction. Therefore $F^{L\prime\prime}(W) < 0$ in the neighbourhood of $\overline{W}^L$.

The rest of the argument about F^L being also concave besides the neighbourhood of $\overline{W}^L$ follows the standard argument. The proof also implies immediately that $rV^L(W) + \gamma W \leqslant \mu$ for all W if the boundary condition $V^{L\prime\prime}(\overline{W}^L) = L$ is true. This conclusion is used in the following verification theorem.

Verification Theorem: for any incentive compatible contract, define an auxiliary gain process G as

$$G_t = \int_0^t e^{-rs}(dY_s - dI_s) + e^{-rt}V(W_t) \ ,$$

where W_t evolves according to dW_t. By Ito's lemma

$$\begin{aligned} e^{rt}G_t &= \left(\mu + \gamma W_t V'(W_t) + \frac{1}{2}\beta_t^2\sigma^2 V''(W_t) - rV(W_t)\right) dt \\ &\quad -(1 + V'(W_t))dI_t + (1 + \beta_t V(W_t))\sigma dZ_t \ . \end{aligned}$$

The first two terms are negative and therefore G_t is a supermartingale. Now evaluating the principal's payoff for this contract

$$E\left[\int_0^\tau e^{-rs}(dY_s - dI_s) + e^{-r\tau}L\right] =$$

$$E\left(G_{t\hat{}\tau}\right) + e^{-r\tau}E\left[1_{\{t\leqslant\tau\}}\left(E_t\left(\int_t^\tau e^{-r(s-t)}(dY_s - dI_s) + e^{-r(\tau-t)}L\right) - V(W_t)\right)\right] \ .$$

First, $E(G_{t\hat{}\tau}) \leqslant G_0$ since G_t is a supermartingale. Then, $E_t(\int_t^\tau e^{-rs}(dY_s - dI_s) + e^{-rt}L) \leqslant \frac{\mu}{r} - W_t$, since by the argument above, $rV(W) + \gamma W \leqslant \mu$ for all W . Letting

$t \to \infty$ implies that

$$E\left(\int_0^\tau e^{-rs}(dY_s - dI_s) + e^{-r\tau}L\right) \leqslant V_0(W)\ .$$

□

Proof of Corollary 2.2.1:The relationship between $V(W)$ and $V^L(W)$ is fairly straightforward: if $V^L(W) > V(W)$ instead, then $V(W)$ cannot be the optimal value function for the principal since the contracting space with the commitment constraint is a strict subset of the contracting space without the constraint.The relationship between $\overline{W}$ and $\overline{W}^L$ follows that $rV(\overline{W}) + \gamma\overline{W} = \mu$ and $rV(\overline{W}^L) + \gamma\overline{W}^L \leqslant \mu$ and the inequality is strict whenever $V(\overline{W}) < L$□

Proof of Proposition 2.3.1: Since the high volatility state is assumed a absorbing state, the value function in such state follows directly from Proposition 2.2.1. The optimality conditions for the low volatility state can be proved in a very similar manner as that in Proposition 2.2.1. Differentiate the corresponding social benefit function with respect to W, substituting in the boundary conditions and evaluating the equation at the payment boundary $\overline{W}_L$ implies

$$F_l'''(\overline{W}_l) = \frac{(\gamma - r) + \left(\gamma\overline{W}_l - \pi_l\delta_l(\overline{W}_L)\right)F_l''(\overline{W}_l)}{\frac{1}{2}\lambda^2\sigma_l^2}\ ,$$

where $F_l''(W_l)$ is given by

$$F_l''(\overline{W}_l) = \frac{rF_l(\overline{W}_l) + (\gamma - r)\overline{W}_l - \mu + \pi_l\left(F_h(\overline{W}_l + \delta_l(\overline{W}_l)) - F_l(\overline{W}_l)\right)}{\frac{1}{2}\lambda^2\sigma_l^2}\ .$$

Piskorski and Tchistyi (2010) show the optimality conditions for the full commitment case. Under limited commitment, if the commitment constraint is not binding in the low volatility state, the proof is identical to theirs. Now suppose it is binding, which implies that it must also be binding in the high volatility state. Given the fact that $V_l^{L\prime}(\overline{W}_l^L) = V_h^{L\prime}(\overline{W}_h^L) = -1$, the slope matching procedure that pins down δ implies $\delta_l^L(\overline{W}_l^L) = \overline{W}_h^L - \overline{W}_l^L$, Given that $rF_l^L(\overline{W}_l^L) + \gamma\overline{W}_l^L < \mu$ from Corollary 1, if $\overline{W}_h^L < \overline{W}_l^L$, then $\delta_l^L(\overline{W}_l^L) < 0$, and $F_l^{L\prime\prime\prime}(\overline{W}_l^L) > 0$. If $\overline{W}_h^L > \overline{W}_l^L$, then $\gamma\overline{W}_l^L - \pi_l\delta_l^L(\overline{W}_l^L) > 0$ as $\pi_l < \frac{\gamma\overline{W}_l^L}{\delta_l^L(\overline{W}_l^L)}$. Since $\delta_l^L(\overline{W}_l^L) < \overline{W}_h^L < \frac{\mu - rL}{\gamma}$, $\gamma\overline{W}_l^L - \pi_l\delta_l^L(\overline{W}_l^L) > 0$ as long as $\pi_l < \frac{\overline{W}_l^L}{\mu - rL}$. Note that for a non trivial contract, $\overline{W}_l^L > R = 0$, there is always π_l small enough such that $\pi_l < \frac{R}{\mu - rL}$ is satisfied. The subsequent verification is similar to that used in Piskorski and Tchistyi (2010) thus is omitted here.□

Proof of Corollary 2.3.1: From Corollary 1 and Proposition 2, $V_h(W) < V_l(W)$ and $V_h^L(W) < V_l^L(W)$for all $W > R$. For the full commitment contract, $V_l''(\overline{W}_l) = V_h''(\overline{W}_h) = 0$ implies $rV_s(\overline{W}_s) + \gamma\overline{W}_s$ for $s = l, h$, then Corollary 1 implies $\overline{W}_l < \overline{W}_h$. For the limited commitment contract, $V_l^L(\overline{W}_l^L) = V_h^L(\overline{W}_h^L) = L$ and $V_s'(W) < 0$ near the payment boundary implies $\overline{W}_l^L > \overline{W}_h^L$□

Proof of Proposition 2.3.2: This proposition is proved in two steps. First, I show that both V_l' and V_h' are convex functions at the payment boundary. This conclusion utilizes the concavity of the value function which is true for both full commitment and limited commitment so only the former is shown. Differentiate the principal's HJB equation with respect to W, and substituting in $V_h'(W + \delta_l(W)) = V_h'(W)$, a condition that is always satisfied around the neighbourhood of the payment boundary because $V_s'(\overline{W}) = -1$ regardless of state and contract type. This yields

$$rV_s'(W) = (\gamma W - \pi_s\delta_s(W))\, V_s''(W) + \frac{1}{2}\lambda^2\sigma_s^2 V_s'''(W) + (\gamma - \pi_s\delta_s'(W))\, V_s'(W)\,.$$

Evaluating this at the payment boundary in the high volatility state yields

$$V_h'''(\overline{W}_h) = \frac{(\gamma - r) - \gamma \overline{W}_h V_h''(\overline{W}_h)}{\frac{1}{2}\lambda^2\sigma_h^2} > 0 \ ,$$

since $V_h''(\overline{W}_h) \leqslant 0$. Similarly,

$$V_l'''(\overline{W}_l) = \frac{(\gamma - r) - \gamma \overline{W}_l V_l''(\overline{W}_l) + \pi_l X(\overline{W}_l)}{\frac{1}{2}\lambda^2\sigma_l^2} \ ,$$

where

$$X(\overline{W}_l) = \delta_l'(\overline{W}_l)V_l'(\overline{W}_l) + \delta_l(\overline{W}_l)V_l''(\overline{W}_l) \ .$$

Letting $\pi_L \to 0$ yields

$$V_l'''(\overline{W}_l) = \frac{(\gamma - r) - \gamma \overline{W}_l V_l''(\overline{W}_l)}{\frac{1}{2}\lambda^2\sigma_l^2} > 0 \ .$$

Therefore $V_l'''(\overline{W}_l) > 0$ for small enough π_l, and both V_l' and V_h' are convex functions at the payment boundary. The same argument applies for the limited commitment contract.

The second step compares the variation of δ near the payment boundary. Consider the full commitment contract: by Corollary 1, $V_h(W) < V_l(W)$ and $V_h(R) = V_l(R) = L$ implies $V_h'(R) < V_l'(R)$. Since $V_h'(\overline{W}_h) = V_l'(\overline{W}_l) = -1$, there must exist $\widehat{W}$ such that $V_h'(\widehat{W}) = V_l'(\widehat{W})$. Moreover, $\overline{W}_h > \overline{W}_l$ by Corollary 2.3.1 and, because V_s' are convex functions, there is a unique $\widehat{W}$ after which $\delta_l(W) > 0$ for all $W > \widehat{W}$.

For the limited commitment contract, $V_h^{L\prime}(R) < V_l^{L\prime}(R)$, $V_h^{L\prime}(\overline{W}_h^L) = V_l^{L\prime}(\overline{W}_l^L) = -1$ and $\overline{W}_h^L < \overline{W}_l^L$ by Corollary 2 implies there exists $\widehat{W}^L$ such that $\delta_l^L(W) > 0$ for all $W > \widehat{W}^L$ Let $\widehat{W}$ be the largest between the two cut-offs for the full and limited commitment contract, and note that $\widehat{W} < \overline{W}_l^L$ since $\delta_l^L(\overline{W}_l^L) < 0$ and $\delta_l(\overline{W}_l) > 0$ proves this proposition.$\square$

Proof of Corollary 2.3.2

Define $\Delta(W) = \left(\overline{W}_h - (W + \delta_l(W))\right) - \left(\overline{W}_l - W\right)$ as the difference between the distances to the payment boundary before and after the uncertainty shock for the full commitment contract, and $\Delta^L(W)$ as the same distance but for the limited commitment contract. Then $\Delta^L(W) - \Delta(W) = \left(\overline{W}_h^L - \overline{W}_h\right) - \left(\overline{W}_l^L - \overline{W}_l\right) - \left(\delta_l^L(W) - \delta_l(W)\right)$. For small π_l, $\overline{W}_l^L - \overline{W}_l$ is small. Therefore $\Delta^L(W) - \Delta(W) < 0$ as long as $\overline{W}_h^L - \delta_l^L(W) < \overline{W}_h - \delta_L(W)$. Notice that $\overline{W}_h - \overline{W}_h^L = \delta_L(\overline{W}_l) - \delta_l^L(\overline{W}_l^L)$, and $\delta_l'(\overline{W}_l) > 0$ while $\delta_l^{L\prime}(\overline{W}_l^L) < 0$ by Proposition 3. Therefore $\delta_l(W) - \delta_l^L(W) < \delta_l(\overline{W}_l) - \delta_l^L(\overline{W}_l^L) = \overline{W}_h - \overline{W}_h^L$ for any $W > \widehat{W}$. That is, $\Delta^L(W) - \Delta(W) < 0$ for all $W > \widehat{W}$.□

Proof of Proposition 2.3.3: Following Cox and Miller (1977), the transition density of the process W in the high variance state given initial value W_{t^+} follows the Kolmogorov forward equation:

$$\frac{\partial}{\partial t} f(t, W; W_{t^+}) = \frac{1}{2}\frac{\partial^2}{\partial W^2}\left[\lambda^2\sigma_h^2 f(t, W; W_{t^+})\right] - \frac{\partial}{\partial W}\left[\gamma W f(t, W; W_{t^+})\right] ,$$

subject to the boundary conditions

$$f(t, 0; W_{t^+}) = 0$$

$$\frac{1}{2}\frac{\partial}{\partial W}\left[\lambda^2\sigma_h^2 f(t, W; W_{t^+})\right]|_{W=\overline{W}_h} - \gamma\overline{W}_h f(t, \overline{W}_h; W_{t^+}) = 0 ,$$

where f is a density function conditional on $W_{t^+} = W$.

Define $\sigma^2 = \lambda^2\sigma_h^2$ as the overall variance of the W process. Let f_γ be the solution to this boundary value problem for a particular γ. According to Ward and Glynn (2003), when γ is closer to zero, f_γ can be approximated by

$$f_\gamma(t, W; W_{t^+}) = k(\gamma)g(t, W; W_{t^+}) + o(\gamma) , \tag{A.1}$$

where $k(\gamma) = \left(1 - \frac{\gamma}{2\sigma^2} W_{t^+}^2 + \frac{\gamma}{2\sigma^2} W^2 + \frac{\gamma}{2} t\right)$ and g is the corresponding transition density function for the same process but with $\gamma = 0$.

Now the problem becomes a Brownian motion between an absorbing and a reflecting barrier. In particular, $g(t, W; W_{t^+})$ satisfies the differential equation:

$$\frac{\partial}{\partial t} g(t, W; W_{t^+}) = \frac{1}{2} \frac{\partial^2}{\partial W^2} \left[\sigma^2 g(t, W; W_{t^+})\right] \ ,$$

subject to boundary conditions $g(t, R; W_{t^+}) = 0$, $\frac{1}{2}\sigma^2 \frac{\partial}{\partial W} [g(t, W; W_{t^+})]|_{W = \overline{W}_h} = 0$.

The solution to this problem has been derived by Schwarz (1992) as

$$g(W, t) = \sum_{n=1}^{\infty} A_n \exp\left(-\alpha_n^2 \frac{1}{2} \sigma^2 t\right) \cos\left(\alpha_n W\right) \ ,$$

where $\alpha_n = \frac{(2n-1)\pi}{2\overline{W}_h}$ and $A_n = \frac{\cos\left(\alpha_n W_{t^+}\right)}{\overline{W}_h}$.

Substituting this into the approximation function (A.1) yields $f(W, t)$ which can be used in the definition of the expected local time at the payment boundary

$$E\left[L_h(T; W_{t^+})\right] = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T dt \int_{\overline{W}_h - \varepsilon}^{\overline{W}_h + \varepsilon} f(t, W; W_{t^+}) dW$$

Fixed some $W_{t^+} < \overline{W}_h^L$, Let

$$E[\mathcal{L}_h(T; W_{t^+})] \equiv E[\mathcal{L}_h(T; W_{t^+}) | \overline{W}_h]$$

be the expected local time given the full commitment value functions and payment boundaries, and define

$$E^L[\mathcal{L}_h(T; W_{t^+})] \equiv E[\mathcal{L}_h(T; W_{t^+}) | \overline{W}_h^L]$$

be the expectation of local time at the payment boundary under the limited commitment contract. First, $\frac{\partial}{\partial T} E[\mathcal{L}_h(T; W_{t^+})]|_{T=0} > 0$, that is, the expected time spend at one point is longer whenever the time interval is longer, in particular when the time interval expands a small amount from 0. Secondly, such derivative is larger for smaller $\overline{W}_h$ because for a fixed W, $f(W,t)$ is decreasing in $\overline{W}_h$. The effect of expanding the time interval is bigger, the shorter distance between W_{t^+} and the reflecting boundary is. Note that in the case of $\sigma \gg \gamma$, the approximation adjustment term $h(\gamma)$ is close to one if W and W_{t^+} are near each other, this implies the most precise approximation is around the payment boundary, exactly the target of the analysis given here.

From Corollary 2, $\overline{W}_h > \overline{W}_h^L$, and $E^L[\mathcal{L}_h(0;W_{t^+})] = E[\mathcal{L}_h(0;W_{t^+})] = 0$ implies

$$E^L\left[\mathcal{L}_h(T;W_{t^+})\right] > E\left[\mathcal{L}_h(T;W_{t^+})\right]\ ,\ as\ T \to 0\ .$$

The expected local time grows faster for closer reflecting boundary near $T = 0$. Also

$$E^L\left[\mathcal{L}_h(T;W_{t^+})\right] < E\left[\mathcal{L}_h(T;W_{t^+})\right]\ ,\ as\ T \to \infty\ ,$$

which implies there is some $\widehat{T}$ such that

$$E^L\left[\mathcal{L}_h(\widehat{T};W_{t^+})\right] = E\left[\mathcal{L}_h(\widehat{T};W_{t^+})\right]\ ,$$

and

$$E^L\left[\mathcal{L}_h(T;W_{t^+})\right] > E\left[\mathcal{L}_h(T;W_{t^+})\right]\ , for\ all\ 0 < T < \widehat{T}\ .$$

Finally, notice that given $\overline{W}_h$, $E\left[L_h(T;W_{t^+})\right]$ is decreasing in W_{t^+}, that is, the further W_{t^+} is from the reflecting barrier, the less time it spends there within a certain time. Therefore $E^L\left[\mathcal{L}_h(T;W_{t^+}^L)\right] > E\left[\mathcal{L}_h(T;W_{t^+})\right]$ as long as $\overline{W}_h^L - W_{t^+}^L < \overline{W}_h - W_{t^+}$. By Corollary 3 $\overline{W}_l^L - W_{t^+}^L < \overline{W}_h - W_{t^+}$, if $W_{t^-} > \widehat{W}$, therefore $E^L\left[\mathcal{L}_h(T;W_{t^-})\right] > E\left[\mathcal{L}_h(T;W_{t^-})\right]$ for all $0 < T < \widehat{T}$ as long as $W_{t^-} > \widehat{W}$□

Proof of Proposition 2.3.4: Consider the process of W in the high volatility state with initial position W_{t^+}. Let N be the number of times W reaches the reflecting boundary $\overline{W}_h$ before it is stopped. Then

$$E[\tau] = \sum_{i=0}^{\infty} E[\tau, N = i] \ .$$

First consider $N \geqslant 1$, if W reaches $\overline{W}_h$ at least once before it is stopped, then starting from $\overline{W}_h$, the expected stopping time is smaller whenever $\overline{W}_h - R$ is a shorter interval. Next consider the case $M = 0$, the expected stopping time is smaller whenever W_{t^+} is closer to R. Finally, the average speed of growth of W, γW, is slower for smaller W. From Corollary 2.3.1 and 2.3.2 it can be concluded that $E^L[\tau_h] < E[\tau_h]$because $\overline{W}_h^L < \overline{W}_h$ and $W_{t^+}^L < W_{t^+}$ for the same W_{t^-}.

Same comparison can be conducted between $E^L[\tau_h]$ and $E^L[\tau_l]$. The expected stopping time is smaller when $\overline{W}$ and the initial W is closer to R, and when σ is larger.

The exact value of $E[\tau]$ is difficult to compute due to the irregular process W follows. However, when γ is small, the same approximation method used in the proof of Proposition 2.3.3 can be applied here as well. The problem thus becomes a standard absorbing time question for a Brownian motion between an absorbing and a reflecting barrier, whose solution is given by Cox and Miller (1977) as

$$E[\tau] = \frac{W_{t^+}(2\overline{W}_h - W_{t^+})}{\sigma^2}$$

This solution confirms that $E[\tau]$ is positively related to $\overline{W}_h$ and W_{t^+} while negatively related to σ. Since $\overline{W}_h^L < \overline{W}_l^L < \overline{W}_h$, $W_{t^+}^L < W_{t^-}^L < W_{t^+}$, and $\sigma_h > \sigma_l$, $E^L[\tau_h]$ must be the smallest compare to $E[\tau_h]$ and $E^L[\tau_l]$ □

Proof of Proposition 2.4.1:

Without loss of generality, assume that the interest rate of the credit line is γ. Begin with the high volatility state σ_h, Then the credit line balance evolves according to:

$$dM_t = \gamma M_t dt + xdt + dDiv_t - dY_t \tag{A.2}$$

where Div_t represents the cumulative dividends paid by the firm and x is the consol bond rate. Using the fact that $x = rD_t$ and substituting that into equation (A.2) implies:

$$dWt = -\lambda dM_t = -\lambda\gamma Mtdt - \lambda xdt - \lambda dDiv_t + \lambda dY_t$$

$$= \gamma W_t dt - \lambda dI_t + \lambda(dY_t - \mu dt)$$

satisfying incentive compatibility. The argument for state σ_L can be made analogously subject to a jump δ_L, whose value is pinned by the matching first order derivatives procedure□

Proof of Proposition 2.4.2:

This Proposition is a natural extension given the proofs of Proposition 1 and 2. The exact conditions under which shirking forever is optimal can be found in Zhu (2013)□

A.2 Proofs for Chapter 3

In this appendix I first detail the system of equations that characterize land price P and firm's portfolio choice α.

Financial Integration: in this equilibrium, firms choose optimal portfolio according to

$$\alpha = \frac{\mu P - rP^2}{\beta\sigma^2} \tag{A.3}$$

firms' net worth follows a geometric Brownian motion:

$$\frac{dW_t}{W_t} = \left[\frac{(\mu - rP)^2}{\beta\sigma^2} + r\right] dt + \frac{\frac{\mu}{P} - r}{\beta(\frac{\sigma}{P})} dZ_t$$

Define $\lambda = \frac{\mu - rP}{\sigma}$ as the Sharp ratio from investing in land, obviously λ must be larger than 0 in equilibrium, otherwise firms will hold risk free bonds only. Substituting λ into the dynamics of firm net worth yields

$$\frac{dW_t}{W_t} = \left[\frac{\lambda^2}{\beta} + r\right] dt + \frac{\lambda}{\beta} dZ_t \tag{A.4}$$

Combine (A.3) and (A.4) into firm's value function yields

$$\rho V = \left(\frac{\lambda^2}{2\beta} + r\right) W_t V'(W_t)$$

The solution is $V(W_t) = C_1 W^\gamma$, where $\gamma = \rho\left(\frac{\lambda^2}{2\beta} + r\right)^{-1} = \frac{2\rho\beta}{\lambda^2 + 2r\beta}$. The coefficient C_1 and payout boundary can be pinned down by matching the refinancing boundary condition: $V'(\underline{W}) = \gamma C_1 \underline{W}^{\gamma-1} = 1 + \zeta$ and the payout boundary condition $V'(\overline{W}) = \gamma C_1 \overline{W}^{\gamma-1} = 1$.

$$C_1 = \frac{1+\zeta}{\gamma} \underline{W}^{1-\gamma}$$

$$\overline{W} = (1+\xi)^{\frac{1}{1-\gamma}} \underline{W}$$

For the refinancing boundary $\underline{W}$, Substituting $V(W) = C_1 W^\gamma$ into boundary condition (3.6) yields

$$C_1^\gamma \underline{W} = \phi + (1+\zeta)\underline{W}$$

and

$$\gamma C_1^{\gamma-1}\underline{W} = 1+\zeta$$

together one can solve for $\underline{W}$

$$\underline{W} = \frac{\gamma\phi}{(1-\gamma)(1+\zeta)}$$

Notice that $\beta \equiv -\frac{WV''(W_t)}{V'(W_t)} = 1-\gamma$, which implies the following relationship

$$2r\beta^2 + \left[2(\rho-r)+\lambda^2\right]\beta - \lambda^2 = 0 \tag{A.5}$$

Define $\Phi(\beta,\lambda^2)$ as the left hand side (LHS) quadratic function of this equation. $\Phi=0$ thus determines β given λ, or equivalently, given P. The power of the value function γ is given by $\gamma = 1-\beta$.

The geometric Brownian motion $\frac{dW_t}{W_t} = \mu_W dt + \sigma_W dZ_t$ with two reflecting barriers $\overline{W}$ and $\underline{W}$ has a stationary distribution. Define $\eta = \frac{2\mu_W}{\sigma_W^2}$, the stationary distribution is characterized by the density function:

$$f(W) = \frac{\eta-1}{\overline{W}^{\eta-1}-\underline{W}^{\eta-1}}W^{\eta-2} \tag{A.6}$$

Finally, P^I is the solution to the market clearing condition

$$\alpha E(W) = P \tag{A.7}$$

where

$$E(W) = \frac{(\eta-1)}{\eta\left(\overline{W}^{\eta-1}-\underline{W}^{\eta-1}\right)}\left(\overline{W}^{\eta}-\underline{W}^{\eta}\right)$$

Combining equations (A.3), (A.4), (A.5), and (A.7) solves for P^I.

Autarky: in this equilibrium, firms invest all their net worth in land, thus $dW_t = \frac{1}{P_t}W_t dA_t$ when W_t is in between the refinancing and payout boundary. The value function satisfies

$$\rho V = \left(\frac{\mu}{P} - \frac{\beta\sigma^2}{2P^2} \right) W_t V'(W_t)$$

Therefore $V(W_t) = C_1 W^{\gamma}$, where $\gamma = \frac{2\rho P^2}{2\mu P - \sigma^2 \beta} \in (0,1)$ and $\gamma = 1 - \beta$ implies:

$$\sigma^2 \beta^2 - \left[2\mu P + \sigma^2 \right] \beta + 2P(\mu - \rho P) = 0 \tag{A.8}$$

The dynamics of firm net worth follow another geometric Brownian motion:

$$\frac{dW_t}{W_t} = \frac{\mu}{P} dt + \frac{\sigma}{P} dZ_t \tag{A.9}$$

There is a stationary distribution for W_t whose density function is given by (A.6). The density function implies the aggregate net worth $E(W)$ which pins down P^A from the solution to the market clearing condition:

$$E(W) = P \tag{A.10}$$

Combining equations (A.8), (A.9), and (A.10) solves for P^A.

Proof of Lemma 3.2.1:

Take (A.5) as a function of β, it is obvious that the function has one negative root and one positive root. The left hand side is smaller that 0 when $\beta = 0$ and larger than 0 when $\beta = 1$, hence there must be a unique solution β such that $0 < \beta < 1$, which implies $0 < \gamma < 1$.

For the autarky equilibrium, notice that $\mu - \rho P^A > 0$, since $\frac{\mu}{\rho}$ is the price of capital when $\sigma = 0$, i.e when investment on capital is risk free. The left hand side of (A.8) is smaller than 0 when $\beta = 1$, therefore (A.8) has one root larger than 1 and one smaller than 1, that is, there is a unique β such that $0 < \beta < 1$.

Proof of Proposition 3.2.1:

To show the relationship between P^I and P^A, I start by showing that P^A coincide with the solution of P^I when $\alpha = 1$. Next, I show that P^I is an increasing function of α. By limited enforcement, $\alpha < 1$ in equilibrium under financial integration, therefore it must be that $P^I > P^A$.

Using equation (A.4), the geometric Brownian motion of firm net worth W_t satisfies:

$$\mu_W = \frac{\lambda^2}{\beta} + r$$

$$\sigma_W = \frac{\lambda}{\beta}$$

plus the lower boundary $\underline{W}$ and endogenous upper boundary $\overline{W}$ pinned down by matching the boundary conditions. Notice that $\overline{W}$ is proportional to $\underline{W}$, which is not surprising given the property of power functions. This reduces the ratio of $\overline{W}^\eta - \underline{W}^\eta$ and $\overline{W}^{\eta-1} - \underline{W}^{\eta-1}$ into a linear function of $\underline{W}$. Using the relationship between β and λ derived in (A.5), standard algebra yields that the market clearing condition can be written as:

$$\frac{\lambda^2 + (r-\rho)}{2\left(\lambda^2 + r\beta\right)}\alpha = \left(\frac{(1+\xi)^{1+\frac{1}{\beta}} - 1}{(1+\xi)^{\frac{1}{\beta}} - 1}\right)\underline{W}$$

Notice that from (A.5), $\lambda^2 = \frac{2r\beta^2 + 2(\rho - r)\beta}{1-\beta}$, substituting this into the simplified market clearing condition yields:

$$\frac{2(r-\rho) - 2r\beta^2 + (1-\rho)(\beta - 1)}{4(r-\rho) - 2r\beta^2 - 2r\beta}\alpha - \left(\frac{(1+\xi)^{1+\frac{1}{\beta}} - 1}{(1+\xi)^{\frac{1}{\beta}} - 1}\right)\underline{W} = 0$$

Define the left hand side of this equation as $\Psi(\beta, \alpha)$, which is now only a function of two variables: firms' effective risk aversion β and their portfolio choice α. The following lemma establishes the link between the autarky equilibrium and the financial integration equilibrium through $\Psi(\beta, \alpha)$:

Lemma A1: P^A is the solution to $\Psi(\beta, 1) = 0$

Proof: Let $\alpha = 1$, then $1 = \frac{\mu P - rP^2}{\beta\sigma^2}$ implies $r = \frac{\mu P - \beta\sigma^2}{P^2}$. Plugging this back into (A.5) implies $2\rho P^2 - 2\mu(1-\beta)P + \sigma^2\beta(1-\beta) = 0$, which, after rearranging terms, is the same as equation (A.8). Substituting equation (A.8) into (A.4) yields $\mu_W = \frac{\mu}{P}$ and $\sigma_W = \frac{\sigma}{P}$ for the geometric Brownian motion of the dynamics of W_t which is identical to dW_t under autarky. Notice that $E(W)$ is defined through η as a function of μ_W and σ_W only, this implies that the market clearing condition (A.7) when $\alpha = 1$ is identical to the market clearing condition for the autarky equilibrium, therefore P^A must equal to P^I when $\alpha = 1$□

Now comparing P^I and P^A is equal to comparing the the different solution of $\Psi(\beta, \alpha) = 0$ in terms of β and α. By the chain rule, $\frac{\partial P}{\partial \alpha} = \frac{\partial P}{\partial \beta} \cdot \frac{\partial \lambda}{\partial \alpha}$. Signs of the two terms can be determined separately in the following two lemmas:

Lemma A2: $\frac{\partial P}{\partial \beta} < 0$

Proof: By the chain rule, $\frac{\partial P}{\partial \beta} = \frac{\partial P}{\partial \lambda} \cdot \frac{\partial \lambda}{\beta}$. The definition of $\lambda = \frac{\mu - rP}{\sigma}$ implies $\frac{\partial P}{\partial \lambda} < 0$. Rearranging the terms of equation (A.5) yields $\lambda^2 = \frac{2(r-\rho) - 2r\beta^2}{\beta - 1}$, and $\frac{\partial \lambda}{\partial \beta} = \frac{1}{2}\lambda \left[\frac{2(\rho - r\beta)}{(\beta-1)^2}\right]^{-\frac{1}{2}} > 0$, since $\rho > r$ and $\beta < 1$. Therefore $\frac{\partial P}{\partial \beta} < 0$.□

Lemma A3: $\frac{\partial \beta}{\partial \alpha} > 0$ under certain parameter conditions

Proof: from $\Psi(\beta, \alpha) = 0$ one can solve α as a function of β:

$$\alpha = 2\left(\frac{(1+\xi)^{1+\frac{1}{\beta}} - 1}{(1+\xi)^{\frac{1}{\beta}} - 1}\right) \frac{4(r-\rho) - 2r\beta^2 - 2r\beta}{2(r-\rho) - 2r\beta^2 + (1-\rho)(\beta - 1)} \underline{W}$$

Taking logarithm on both side of the equation:

$$\ln \alpha = \ln 2\underline{W} + \ln\left((1+\xi)^{1+\frac{1}{\beta}} - 1\right) - \ln\left((1+\xi)^{\frac{1}{\beta}} - 1\right) + \ln(\lambda^2 + r\beta) - \ln(\lambda^2 + (r-\rho))$$

Differentiate this equation with respect to β yields:

$$\frac{\partial \alpha}{\partial \beta} = A + B + C$$

where

$$A = \frac{(1+1/\beta)\phi\,(1+\xi)^{1/\beta}}{(1+\xi)^{1+\frac{1}{\beta}}-1} + \frac{(1/\beta)\,(1+\xi)^{-1/\beta}}{(1+\xi)^{\frac{1}{\beta}}-1}$$

$$B = \frac{L+r}{\lambda^2 + r\beta}$$

$$C = \frac{L}{\lambda^2 + (r-\rho)}$$

and $L \equiv \frac{\partial}{\partial\beta}\left(\lambda^2\right)$

There are three parts that characterize the derivative of α on β: Part A measures the change of payment boundary $\overline{W}$, this term is positive because lower β implies weaker risk aversion which lowers the payment boundary. Part B comes from the change of η due to the drift term μ_W, which is smaller when α is lower; Part C is the wealth effect reflected by the change of η due to the diffusion term σ_W, which is also smaller when α is lower.

Notice that part A is bounded by:

$$\frac{2\phi\,(1+\xi)}{(1+\xi)^2-1} - \frac{1}{(1+\xi)^2-(1+\xi)} = \frac{2\phi\,(1+\xi)^2-1}{(\xi+2)(\xi+1)}$$

Part A reaches this upper bound, defined as $\overline{A}$, when $\beta = 1$. A sufficient condition for $\frac{\partial\beta}{\partial\alpha} > 0$ is therefore:

$$\frac{L(r\beta+\rho-r)-r}{(\lambda^2+r\beta)(\lambda^2+(r-\rho))} \geqslant \overline{A} \tag{A.11}$$

First, $\overline{A}(\lambda^2 + r\beta)(\lambda^2 + (r-\rho)) \leqslant \overline{A}\lambda^2(\lambda^2 + r\beta)$, and $L = \frac{\partial}{\partial\beta}\left(\lambda^2\right) = \frac{2(\rho - r\beta)}{(\beta-1)^2}$, therefore the inequality of (A.11) is true as long as

$$\frac{2(\rho - r\beta)}{(\beta-1)^2}(r\beta+\rho-r) \leqslant \overline{A}\lambda^2(\lambda^2+r\beta) + r$$

$$2(\rho - r\beta)(r\beta + \rho - r) \leqslant \overline{A}\lambda^2(\lambda^2 + r\beta)(\beta - 1)^2 + r$$

$$2(\rho - r\beta)(r\beta + \rho - r) \leqslant \overline{A}[2(r - \rho) - 2r\beta^2 + r\beta(\beta - 1)^2]$$

$$2\rho((r\beta + \rho - r) \leqslant \overline{A}[2(r - \rho) - 4r\beta^2 + r\beta]$$

In every step above, the left hand side is made larger and right hand side made smaller and by using $r\beta \approx 0$. Finally, since $\beta < 1$, the last inequality above implies the following sufficient condition under which $\frac{\partial\beta}{\partial\alpha} > 0$ is satisfied:

$$\frac{3r + (2r - 1)\rho}{\rho - r} \leqslant \overline{A}$$

Combining Lemma A2 and A3 implies $\frac{\partial P}{\partial\alpha} = \frac{\partial P}{\partial\beta} \cdot \frac{\partial\beta}{\partial\alpha} < 0$. By Lemma A1, P^A coincides with the solution where $\alpha = 1$ and $\alpha < 1$ in the financial integration equilibrium, therefore it must be $P^A < P^I$. □

Proposition 1 implies the following corollary

Corollary 3.2.1: $\beta^{In} < \beta^{Au}$

Proof: From (A.5), β which is a increasing function in λ. λ is a decreasing function in P. Therefore given $P_A < P_I$ this implies $\beta^{In} < \beta^{Au}$. This result is intuitive as precautionary savings with risk free asset implies firms are less risk averse. This Corollary basically states that the wealth effect comes from the reduction of effective risk aversion

Proof of Proposition 3.2.2:

Proof: by the definition of savings rate:

$$S^I \equiv \frac{E^I(W)}{\mu + (1 + r(1 - \alpha))E^I(W)}$$

$$S^A \equiv \frac{E^A(W)}{\mu + E^A(W)}$$

Suppose that $S^I < S^A$. Because $E^I(W) > E^A(W)$, it must be that:

$$\frac{E^I(W)}{\mu + (1 + r(1-\alpha))E^I(W)} < \frac{E^A(W)}{\mu + E^A(W)}$$

$$\mu E^I(W) + E^A(W)E^I(W) < \mu E^A(W) + (1 + r(1-\alpha))E^I(W)E^A(W)$$

That is,

$$\mu(E^I(W) - E^A(W)) < r(1-\alpha)E^I(W)E^A(W)$$

dividing both sides by $E^A(W)$ yields

$$\mu(\frac{E^I(W)}{E^A(W)} - 1) < r(1-\alpha)E^I(W)$$

First, $\frac{E^I(W)}{E^A(W)} = \frac{P^I}{\alpha P^A}$, therefore $\frac{\mu P^I}{\alpha P^A} < r(1-\alpha)E^I(W)$, that is,

$$r\alpha(1-\alpha)E^I(W) > \mu$$

or, $E^I(W) > \frac{4\mu}{r}$. Recall that:

$$E(W) = \frac{(\eta - 1)}{\eta\left(\overline{W}^{\eta-1} - \underline{W}^{\eta-1}\right)}\left(\overline{W}^{\eta} - \underline{W}^{\eta}\right) < \frac{(1+\xi)^{1+\frac{1}{\beta}} - 1}{(1+\xi)^{\frac{1}{\beta}} - 1}$$

Therefore:

$$\frac{(1+\xi)^{1+\frac{1}{\beta}} - 1}{(1+\xi)^{\frac{1}{\beta}} - 1} > \frac{4\mu}{r}.$$

The left hand side has a minimum $1 + \zeta$ when $\beta \to 0$, therefore for $S^I < S^A$ a necessary condition is:

$$(1 + \zeta)r > 4\mu \tag{A.12}$$

When condition (A.12) it is violated, $S^I < S^A$ cannot be true. i.e. when Condition 2: $(1 + \zeta)r < 4\mu$ is satisfied, it must be $S^I > S^A$.□

Proof of Proposition 3.3.1:

The value function of the social planner's problem solves:

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu}{P} - r \right) \alpha_t W_t + r W_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \qquad \text{(A.13)}$$

subject to the market clearing condition:

$$\int \alpha_t W_t dF(W) = P$$

The market clearing condition implies $P = P(\alpha)$ as an explicit function of α. The first order condition of HJB equation (A.13) with respect to α is:

$$\left(\frac{\mu}{P} - r \right) W V' - \frac{\sigma^2}{P^2(\alpha)} \alpha W^2 V'' - \frac{\mu \alpha}{P^2} P'(\alpha) W V' + \frac{\alpha^2}{P^3} \sigma^2 P'(\alpha) = 0$$

The first two terms are the first order derivative of V with respect to α taking P as a constant. Setting them equal to zero gives α^I, the solution of α to the decentralized economy where firms take the equilibrium price P as given. The third and fourth terms come from the additional market clearing constraint for solving α^*, that the social planner's problem taking into account the effect of choosing α on the equilibrium price P. Using the same procedure of refinement as Section III, I look for the solution where $V'' = -\frac{\beta V'}{W}$ for some β. Substituting out V'' and simplifying the first order condition implies:

$$\left(\frac{\mu}{P} - r \right) - \frac{\beta \sigma^2}{P^2} \alpha - \frac{\mu \alpha}{P^2} P' + \frac{\beta \alpha^2}{P^3} \sigma^2 P' = 0 \qquad \text{(A.14)}$$

Let $A(\alpha) \equiv \left(\frac{\mu}{P} - r \right) - \frac{\beta \sigma^2}{P^2} \alpha$, equation (A.14) defines P' as a function of α:

$$P'(\alpha)|_{\alpha = \alpha^*} = \frac{A(\alpha^*) P^3}{\alpha^* (\beta^* \alpha^* \sigma^2 - \mu)}$$

Notice that, $A(\alpha^I) = 0$ since α^I is the solution of setting $A(\alpha) = 0$. Suppose that $\alpha^* > \alpha^I$. The proof of Proposition 1 shows that $P' < 0$ under condition 1, implying that $\beta^*\alpha^*\sigma^2 - \mu < 0$. However, from equation (A.8) algebra shows that $\beta^I\alpha^I > \frac{\mu}{\sigma^2}$. According to Lemma A3, $\beta\alpha$ is increasing in α, thus $\beta^*\alpha^* > \beta^I\alpha^I > \frac{\mu}{\sigma^2}$, contradiction. It must be that $\alpha^* < \alpha^I$□

Proof of Proposition 3.3.2:

Suppose the government imposes a tax τ on land. Firms value function now satisfies:

$$\rho V(W_t) = \max_{\alpha_t \leqslant 1} \left[\left(\frac{\mu}{P} - r - \tau \right) \alpha_t W_t + rW_t \right] V'(W_t) + \frac{1}{2} \frac{\alpha_t^2}{P^2} \sigma^2 W_t^2 V''(W_t) \quad \text{(A.15)}$$

with the first order condition being:

$$\left(\frac{\mu}{P} - r - \tau \right) WV' - \frac{\sigma^2}{P^2} \alpha W^2 V'' = 0$$

Imposing that α_t is a constant, this first order condition becomes:

$$\left(\frac{\mu}{P} - r - \tau \right) - \frac{\beta\sigma^2}{P^2} \alpha = 0$$

When $\alpha^I = \alpha^*$, from the proof of Proposition 3 this implies

$$\left(\frac{\mu}{P} - r \right) - \frac{\beta\sigma^2}{P^2} \alpha - \frac{\mu\alpha}{P^2} P' + \frac{\beta\alpha^2}{P^3} \sigma^2 P' = 0$$

Setting $\frac{\tau P^2}{\beta\sigma^2} = \alpha^*(\beta^*\alpha^*\sigma^2 - \mu)$ yields

$$\tau = \frac{\alpha}{P^2} \left(\frac{\beta\alpha}{P} \sigma^2 - \mu P' \right)$$

From the proof of Proposition 3.2.1, $P'(\alpha) < 0$. Therefore $\tau > 0$ is a subsidy.□

Appendix B

Further Discussions

B.1 Recurring States

In the main body of the paper I assume that the transition probability from high to low uncertainty state π_h is zero, that is the crisis state is absorbing. This assumption greatly simplifies the verification of the optimality of the contract provided by Proposition 2, but is unnecessarily for the results of this paper to hold. In this appendix I provide a full characterization of the optimal contract when I relax such assumption. That is, when $\pi_h > 0$ and the economy switches between normal and crisis times stochastically. The following proposition summarizes the result:

Proposition B.1.1. *Suppose $\pi_l > 0$ and $\pi_h > 0$. Let N_t be the total number of state transitions at time t. The agent's continuation utility W_t follows*

$$dW_t = \gamma W_t - dI_t + \lambda(dY_t - \mu dt) + \delta_t(dN_t - \pi_t dt);. \quad \text{(B.1)}$$

The optimal contract is a pair of value functions $V_s(W)$ and payment boundaries

$\overline{W_s}$, $s \in \{l, h\}$ *such that*

$$\begin{aligned} rV_s(W) &= \mu + (\gamma W - \pi_s \delta_s(W))\, V_s'(W) + \frac{1}{2}\lambda^2 \sigma_s^2 V_s''(W) \\ &\quad + \pi_s \left(V_{\hat{s}}(W + \delta_s(W)) - V_s(W) \right) , \end{aligned} \tag{B.2}$$

subject to boundary conditions $V_s(R) = L$; $V_s'(\overline{W}_s) = -1$; *and*

$$V_s''(\overline{W}_s) = 0 ,$$

where $\delta_s(W)$ *follows (2.7) and (2.8). If the principal has only limited commitment, the optimal contract is a pair of value functions* $V_s^L(W)$ *and payment boundaries* $\overline{W}_s^L$, $s \in \{l, h\}$, *such that* $V_s^L(W)$ *satisfies the same system of ODE (B.2) and boundary conditions* $V_s^L(R) = L$; $V_s^{L\prime}(\overline{W}_s^L) = -1$, *and*

$$V_s^{L\prime\prime}(\overline{W}_s^L) = 0, \text{ if } V_s^L(\overline{W}_s^L) \geqslant L ,$$

$$V_s^L(\overline{W}_s^L) = L, \text{ otherwise} .$$

Proof: the proof builds on iteration procedure described in Li (2012). I therefore only sketch the argument here in the interest of space. Consider first the case of full commitment. Applying the martingale method of Sannikov (2008), the agent's continuation utility follows (B.1). Ito's lemma implies that the principal's HJB equation satisfies (2.6). Let $\widetilde{V}_s(W)$ be a solution to (2.6). The concavity of $\widetilde{V}_s(W)$ can be shown using the method similar to Proposition 2. Take $\widetilde{V}_l(W)$ as given, define an auxiliary value function U_h^S as the payoff assuming the principal ceases to provide any incentive to the agent in the high volatility state until the next volatility shock arrives. The concavity of $\widetilde{V}_s(W)$ implies that $\widetilde{V}_h(W) > U_h^S$. Apply the similar argument to $\widetilde{V}_l(W)$ but take $\widetilde{V}_h(W)$ as given, Li (2012) shows that the procedure converges to a pair of function $V_s(W)$ satisfying equation (B.2). Finally, the same procedure also applied to the limited commitment contract as long as $V_s(W)$ remain concave, which is shown in Proposition 1 by replacing the $V_h(W)$ with $\widetilde{V}_h(W)$ in its proof.□

The optimal contract characterized under recurring state is qualitatively identical to the one summarized in Proposition 2 under a one-time shock. In fact, principal's value functions of the contract under recurring states converge to value functions under a one-time shock when $\pi_h \to 0$. Given π_s are assumed to be small numbers the case of a one-time shock provides a good approximation for the general case of recurring states and does not lose any important result.

All the remaining results discussed in the main body regarding the position of payment boundaries, the frequency of cash payment and expected termination time are preserved in the recurring state contract, as long as the parameters are that once the limited commitment constraint is imposed, it is binding in both states. The discussion of "pay-for-luck" can be expanded to not only negative shocks but also positive shocks. The following result can be inferred from Proposition 3: Under the full commitment contract, managers whose accumulated performance is well enough receive less utility when volatility decreases; meanwhile managers under the limited commitment contract receive higher utility. The conclusion for limited commitment contract is consistent with empirical findings of "pay-for-luck" which further reinforce the importance of taking firms' commitment ability into account when understanding compensation under shocks.

Contracts with Limited Commitment Binding in One State Only

Section 3.1 introduced three types of contracts based on when the limited commitment constraint is binding. While the main text focuses on the third types, here I also provide some discussions of the second type: the contract where the limited commitment constraint is binding only in the high volatility state. In general, this type of contract can behave like contracts with either full commitment or those with limited commitment but the commitment constraint is binding in both states, depending on the parameter value σ in each state.

The main goal of this appendix is to establish conditions under which the main Propositions in Section 4 are still valid for the optimal contract when the limited commitment constraint is imposed. The proofs of Proposition 4 reveal that the key variable driving the dynamics of compensation is the distance between W_{t^+} and payment boundary $\overline{W}_h$. This leads to the conjecture that the dynamics of compensation of the type of contract discussed in this appendix section will be similar to the dynamics of the limited commitment contracts described in Section 4 as long as when the commitment constraint is imposed, the agent's continuation utility W_{t^+} is closer to the payment boundary $\overline{W}_h^L$ compared to the full commitment case. Due to the implicit form of the value function, it is analytically difficult to characterize the exact range of parameters under which this conjecture is true. Nevertheless the following proposition gives one sufficient condition for it.

Proposition B.1.2. *If the commitment constraint is binding only in the high variance state, then there exist $\widehat{W}$ such that $W_{t^+}^L - \overline{W}_h^L < W_{t^+} - \overline{W}_h$ for all $W_{t^-} > \widehat{W}$ as long as $\overline{W}_h < \overline{W}_l$.*

Proof: Similar to the proof of Proposition 3 and Corollary 3, consider the full commitment contract first. $V_h(W) < V_L(W)$ implies $\overline{W}_h > \overline{W}_l$ and $V_h'(R) < V_l'(R)$. Since $V_h'(\overline{W}_h) = V_l'(\overline{W}_l) = -1$, there must exist $\widehat{W}$ such that $V_h'(\widehat{W}) = V_l'(\widehat{W})$ and $\delta_l(W) > 0$ for all $W > \widehat{W}$.

Next, if the limited commitment constraint is imposed and binding, $V_h^{L\prime}(R) < V_l^{L\prime}(R)$, $V_h^{L\prime}(\overline{W}_h^L) = V_l^{L\prime}(\overline{W}_l^L) = -1$. If $\overline{W}_h^L < \overline{W}_l^L$, then there exists $\widehat{W}^L$ such that $\delta_h^L(W) > 0$ for all $W > \widehat{W}$. Let $\widehat{W} = \max\left\{\widehat{W}, \widehat{W}^L\right\}$, then $\delta_l'(\overline{W}_l) > 0$ while $\delta_l^{L\prime}(\overline{W}_l^L) < 0$ for all $W > \widehat{W}$. Finally, define $\Delta^L(W_{t^-}) - \Delta(W_{t^-}) = \left(\overline{W}_h^L - \overline{W}_h\right) -$

$\left(\overline{W}_l^L - \overline{W}_l\right) - \left(\delta_l^L(W_{t^-}) - \delta_l(W_{t^-})\right)$. For small π_l, $\overline{W}_l^L - \overline{W}_l$ is small. Notice that $\overline{W}_h - \overline{W}_h^L = \delta_l(\overline{W}_l) - \delta_l^L(\overline{W}_l^L)$, $\delta_l'(\overline{W}_l) > 0$ and $\delta_l^{L\prime}(\overline{W}_l^L) < 0$ implies $\delta_l = (W_{t^-}) - \delta_l^L(W_{t^-}) < \delta_l = (\overline{W}_l) - \delta_l^L(\overline{W}_l^L) = \overline{W}_h = -\overline{W}_h^L$ for any $W_{t^-} > \widehat{W}$. Therefore, $\Delta^L(W_{t^-}) - \Delta(W_{t^-}) < 0$ for all $W_{t^-} > \widehat{W}$.□

Given the sufficient condition above, the rest of the analysis follows exactly the one shown in the main text. Figure C.4 demonstrate the difference between two levels of volatility in the high volatility state. For the same level of σ_l, the relative position of $\overline{W}_l$ and $\overline{W}_h^L$ are similar to the full commitment case when σ_h is moderate, but converge to the case in which the commitment constraint is binding in both states when σ_h becomes high enough.

The finding of this section greatly expands the domain of contracts to which Propositions 4 and 5 can apply. Large bonuses in crisis times could be possible if the abrupt volatility increase is substantial enough that many firms that operate smoothly during normal times suddenly become constrained in the amount they can credibly pledge to pay their managers in the long-run. The greater increase of market risks during the crises, the more severe is this concern. Future research that calibrates or empirically investigates the real scope of this commitment constraint will be helpful in determining the proportion of firms that are subject to limited commitment contracts and firms whose dynamics of bonuses follow the predictions in this paper.

Renegotiation-Proof Contracts

A renegotiation-proof contract requires the slope of the principal's value function to be non-positive. Such condition is ruled out in the main context of this paper because $V^L(R) = V^L(\overline{W}^L) = L$ when the limited commitment constraint binds, hence a non-trivial contract must have a region where the principal's valuation is increasing in the agent's continuation value W. To allow renegotiation-proof contracts I modify the assumption about the principal's commitment ability. I assume that, instead of having the option to liquidate the project any time during the contract, the principal will only withdraw the investment when firm values is below zero. This assumption is similar to the one made by Ai and Li (2013). The corresponding constraint on the payment boundary is now:

$$V(\overline{W}) \geqslant 0 \ .$$

The dynamics of the agent's continuation value under renegotiation-proof contracts follow

$$dW_t = \gamma W_t - dI_t + \lambda(dY_t - \mu dt) + \delta_t(dN_t - \pi_t dt) + dP_t \ .$$

The new term dP_t defines a reflecting termination boundary $\underline{W}$ which satisfies the boundary condition $V(\underline{W}) = L$ and $V'(\underline{W}) = 0$. Termination is stochastic at this boundary, with probability $\frac{dP_t}{\underline{W}-R}$ to account for the extra term on the agent's continuation value and keep the contract incentive compatible.

For the main results of this paper to carry through, it is sufficient to prove the following proposition:

Proposition B.1.3. *Under the renegotiation-proof contract, $\overline{W}_h > \overline{W}_l$ under full commitment and $\overline{W}_h^L < \overline{W}_l^L$ when the constraint is binding in both states.*

Proof: Clearly, Corollary 1 and 2 still apply to renegotiation-proof contracts. Therefore $\overline{W}_h^L < \overline{W}_l^L$ since $V_h(W) < V_l(W)$ and $V(\overline{W}) = 0$ if the commitment constraint is binding in both states. Without the constraint the contract is a standard continuous-time dynamic contract with regime switching, and the argument of the boundary positions can be found in Hoffmann and Pfeil (2010)□

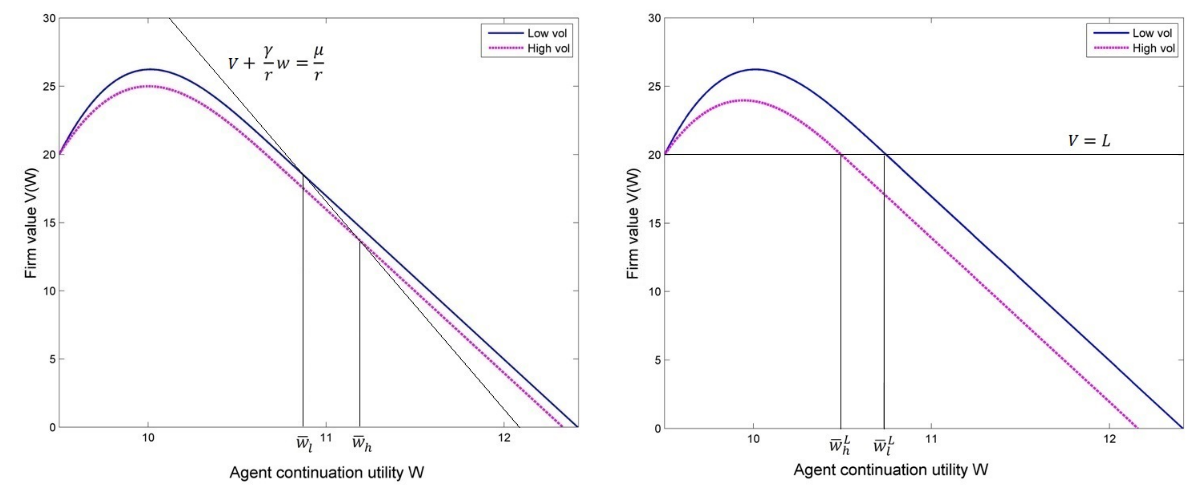
Given the relative positions of the payment boundaries for the full commitment and limited commitment contracts, one can easily see that a statement similar to

Corollary 3 can be made here as well. Figure C.5 shows the value functions for the renegotiation-proof contracts, where both the endogenous renegotiation boundaries as well as the payment boundaries are displayed. The graphs confirms Propositions regarding the relative positions of payment boundaries for both the full and limited commitment contracts. Such conclusions leads to the same dynamics of bonuses payment described in Section 3 and the details are thus omitted here.

Last but not least, renegotiation-proofness is not a necessary feature for the contract to be optimal despite limited commitment. The principal is still able to rule out further renegotiation since the only action she cannot commit to is not to withdraw when firm value is negative. In particular, the principal can commit to the random termination schedule described above, which is crucial in keeping the manager's incentive properly. Further, the assumption of investors withdrawing their investment when firm value drops below zero replaces the earlier assumption of liquidation at any time, and therefore the value of the firm at the termination boundary is still the liquidation value since it is determined by the agent's effective limited liability constraint and the principal is able to commit to termination once that boundary is reached.

Appendix C

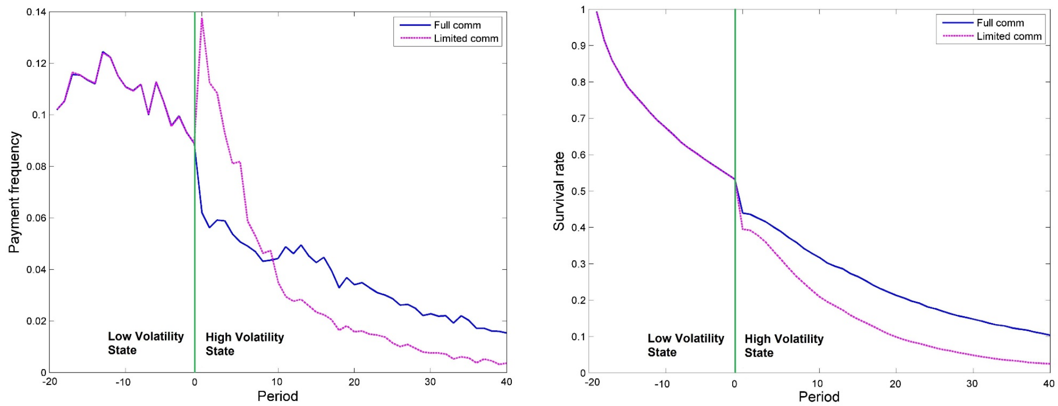
Figures



Panel A. Full Commitment Panel B. Limited Commitment

Figure C.1: **Value Functions for the Optimal Contracts**

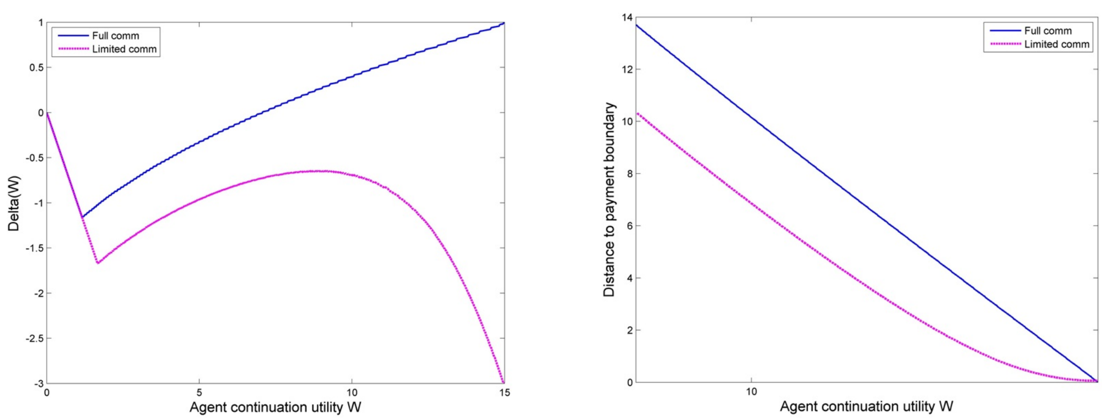
This figure plots firm value functions under regime switching. The full commitment case is shown in the left panel. The limited commitment case is shown in the right panel. Parameter values are $L = 20$, $R = 0$, $\gamma = 0.04$, $r = 0.02$, $\mu = 1$, $\lambda = 0.1$, $\sigma_l = 5.9$, $\sigma_h = 6.5$, $\pi_l = 0.001$, $\sigma_h = 0$



Panel A. Frequency of Payments   Panel B. Survival Rate

FIGURE C.2: **Simulation Results**

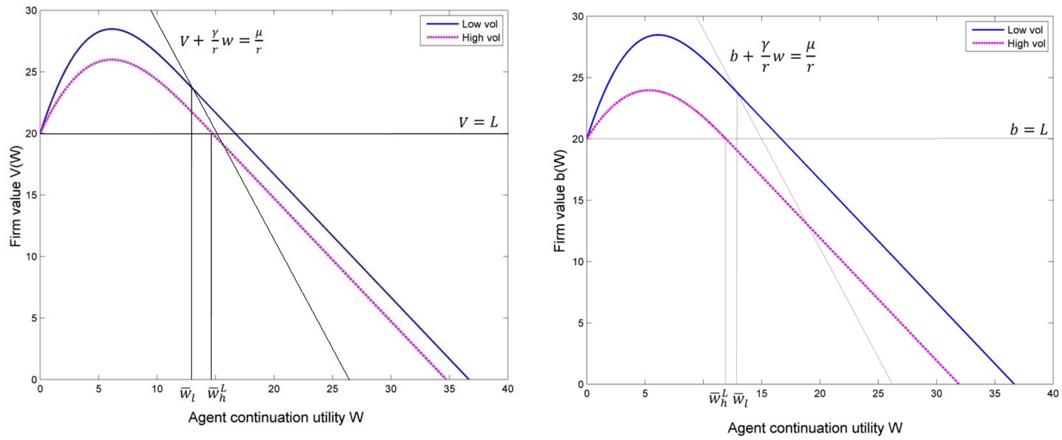
This figure plots the frequency of cash compensation (bonuses) and the fraction of active projects from simulating 5000 paths of cash flows, Parameter values are the same as those in Figure C.2.



Panel A. Size of $\delta_l(W)$ Panel B. Distance to the Payment Boundary

FIGURE C.3: **Allocation of Agent's Continuation Utility**

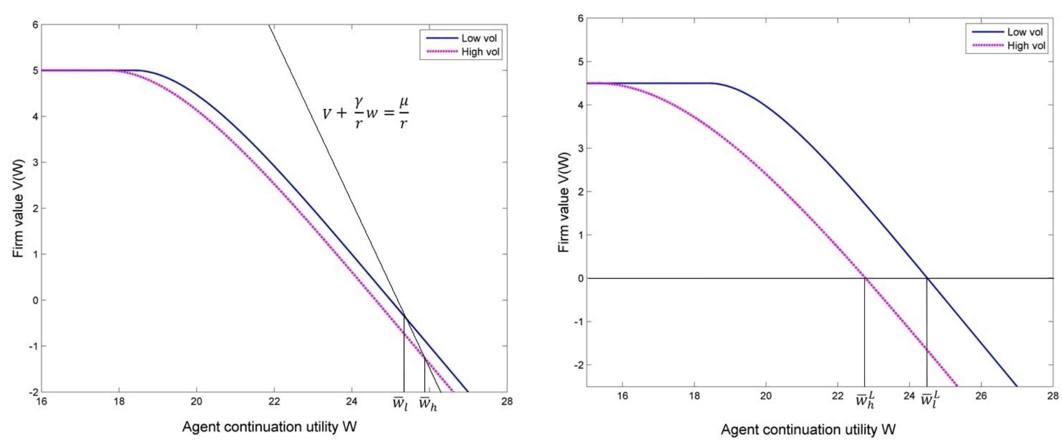
This figure plots the size of δ_l (left panel) and $\overline{W}_h - W_{t^+}$ (right panel), the distance between agent's continuation utility and the payment boundary after the uncertainty shock.



Panel A. Low σ_h Panel B. High σ_h

Figure C.4: **Contracts with the Commitment Constraint Binding in One State Only**

This figure plots firm value functions when the limited commitment constraint is binding only in the high volatility state. Parameter values are the same as those in Figure C.1 except $\sigma_l = 5$ and $\sigma_h = 6$ for the left panel, and $\sigma_h = 6.5$ for the right panel



Panel A. Full Commitment Panel B. Limited Commitment

Figure C.5: **Renegotiation-Proof Contracts**

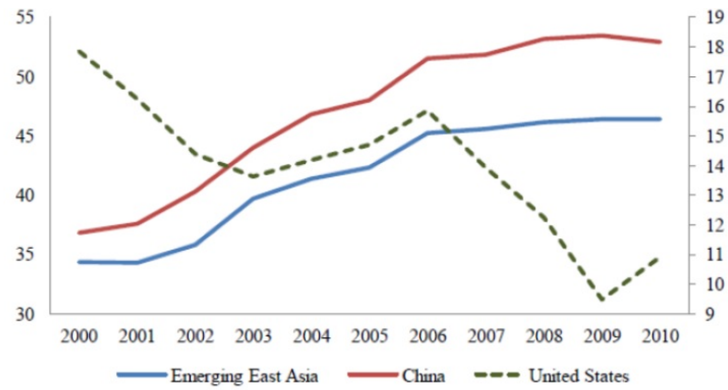


FIGURE C.6: **Gross Savings Rate (in percentage of GDP)**

source: World Bank

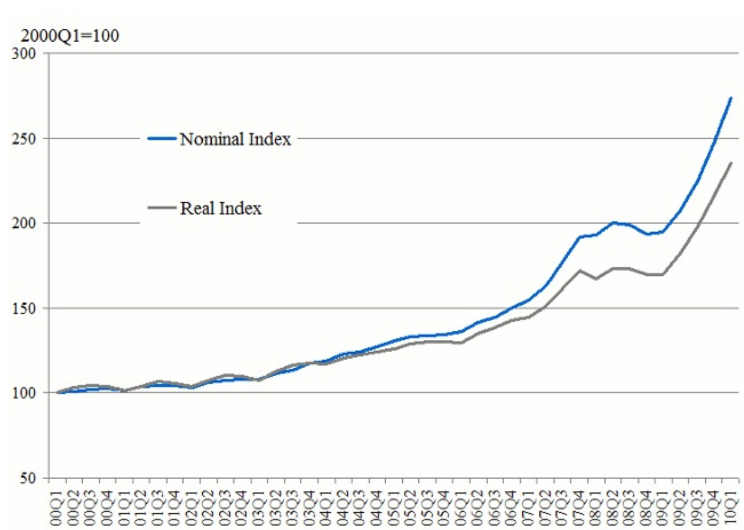


FIGURE C.7: **Housing Index, China**

source: Wu, Gyourko and Deng (2012)

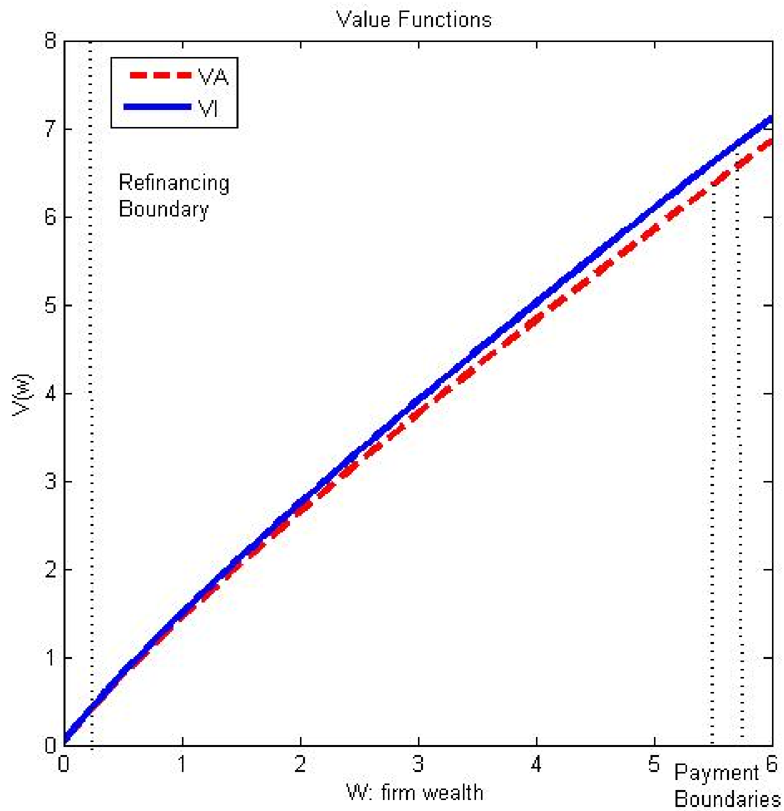
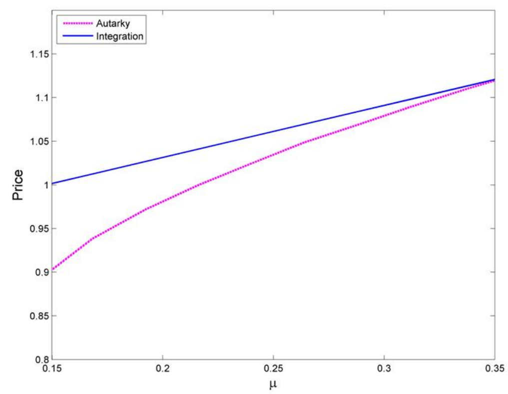


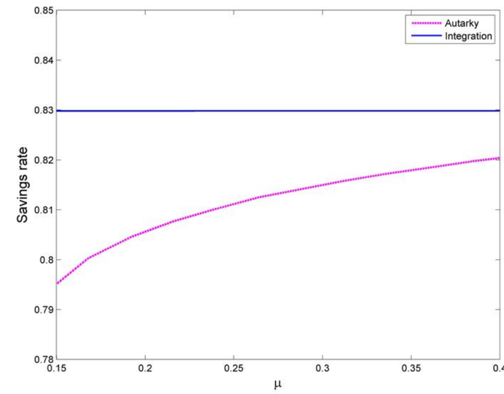
Figure C.8: **Value Function**

This figure shows the value function in autarky (VA) and under financial integration (VI), using parameters in the following table

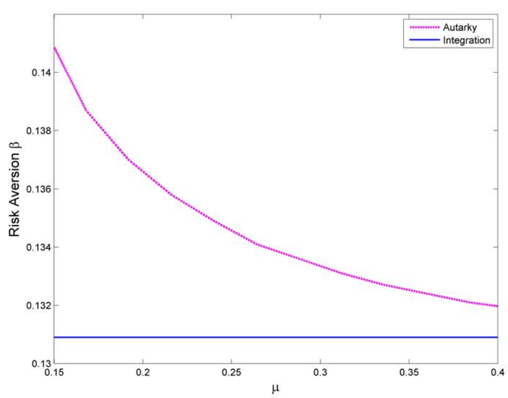
ρ	r	μ	σ	ξ	ϕ
0.15	0.1	0.15	0.5	0.15	0.1



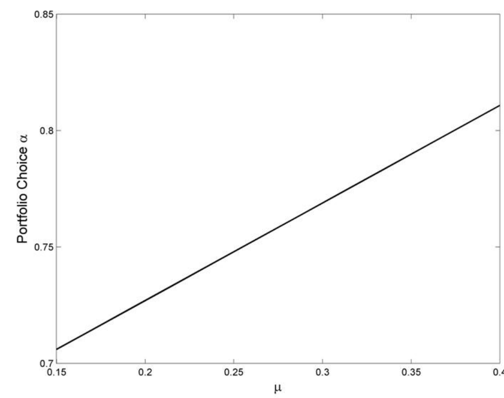
A.Asset Prices



B.Savings Rate



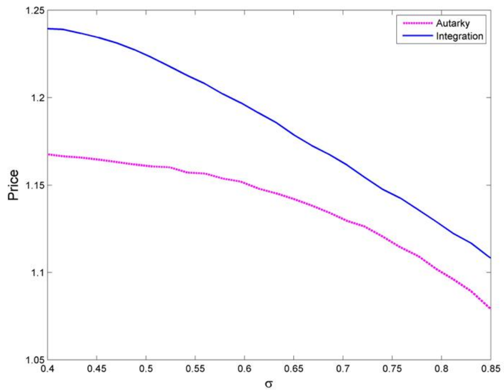
C.Risk Aversion



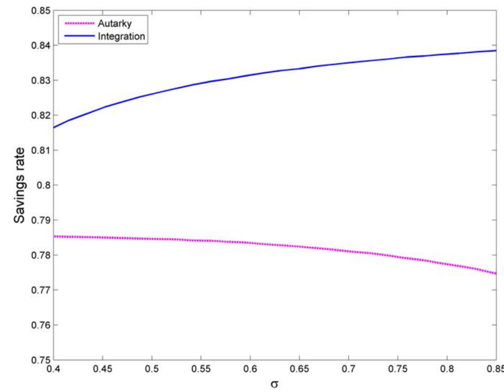
D.Portfolio Choice

FIGURE C.9: **Comparative Statics:** μ

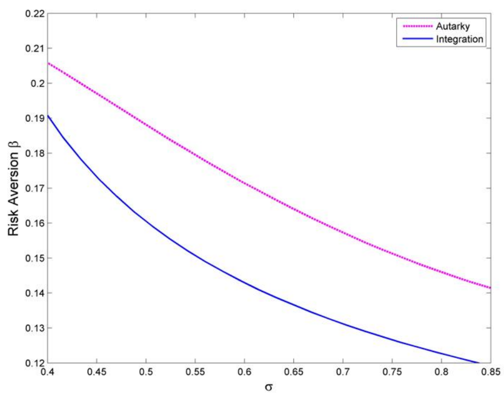
This figure presents the comparative statics of the return on land, μ. The red dotted line indicates the value in autarky while the blue line indicates the value under financial integration



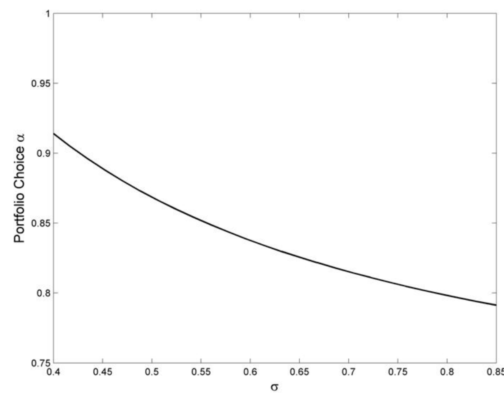
A.Asset Prices



B.Savings Rate



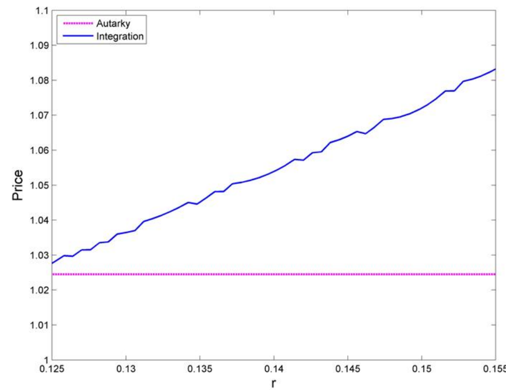
C.Risk Aversion



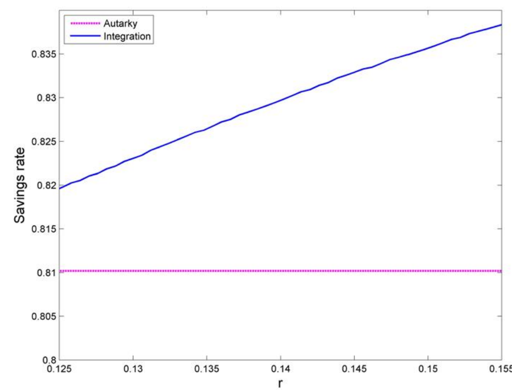
D.Portfolio Choice

Figure C.10: **Comparative Statics: σ**

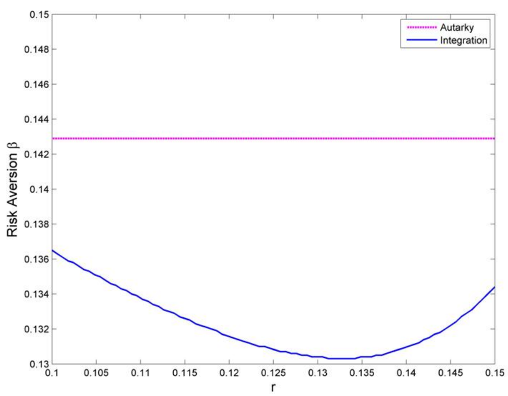
This figure presents the comparative statics of the volatility of return on land, σ. The red dotted line indicates the value in autarky while the blue line indicates the value under financial integration



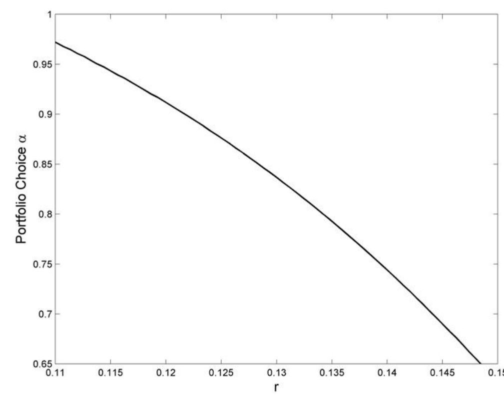
A.Asset Prices



B.Savings Rate



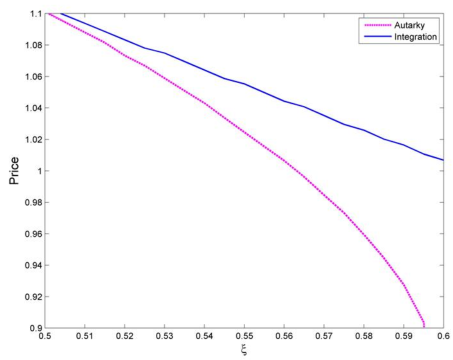
C.Risk Aversion



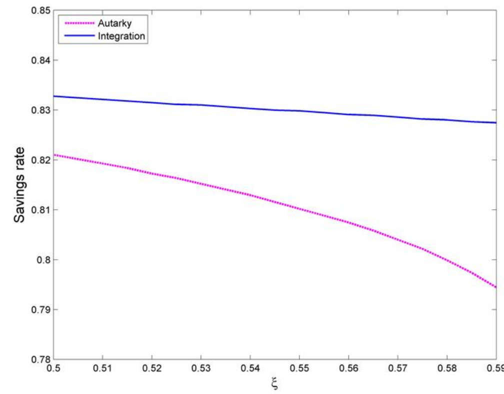
D.Portfolio Choice

FIGURE C.11: **Comparative Statics:** r

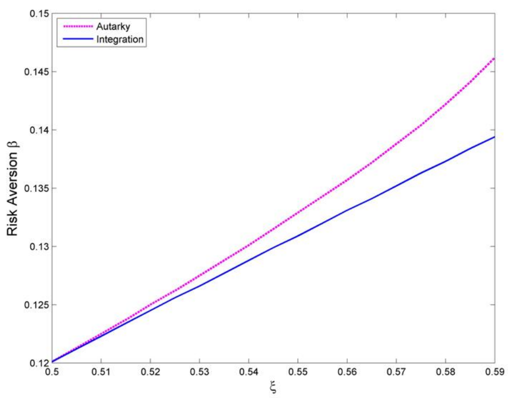
This figure presents the comparative statics of the international risk free interest rate, r. The red dotted line indicates the value in autarky while the blue line indicates the value under financial integration



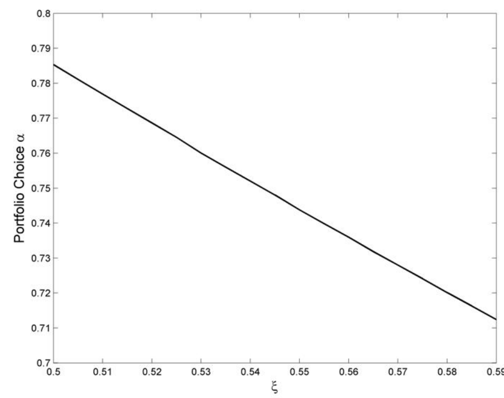
A.Asset Prices



B.Savings Rate



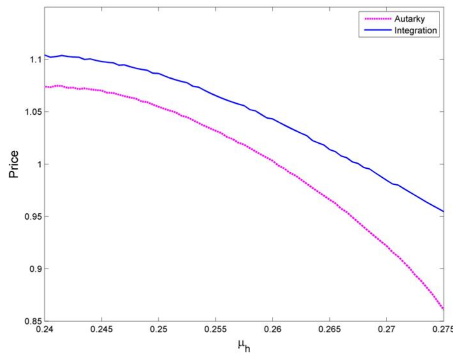
C.Risk Aversion



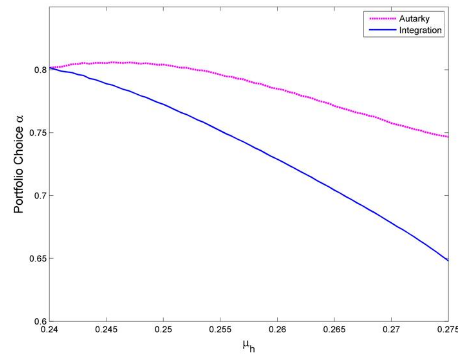
D.Portfolio Choice

FIGURE C.12: **Comparative Statics:** ξ

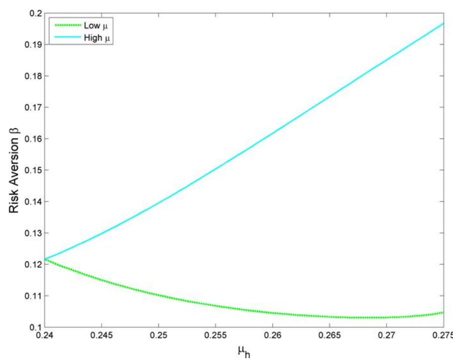
This figure presents the comparative statics of the marginal refinancing cost, ξ. The red dotted line indicates the value in autarky while the blue line indicates the value under financial integration



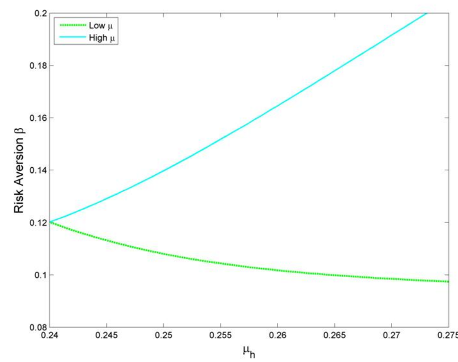
A. Risk Aversion under Autarky



B. Risk Aversion under Financial Integration



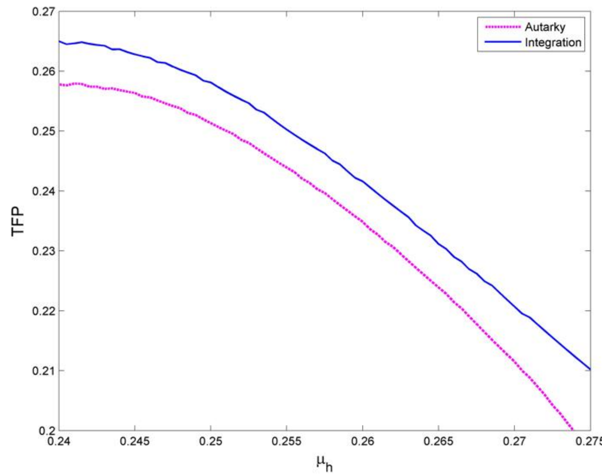
C. Savings Rate under Autarky



D. Savings Rate under Financial Integration

Figure C.13: **Heterogenous Productivity, Asset Prices, and Risk Aversion**

This figure presents the effect of productivity heterogeneity on land prices. Panel A shows the land prices under autarky (red dotted) and financial integration (blue). Panel B shows the portfolio choice α for both low productivity and high productivity firms. Panel C and D shows the degree of risk aversion



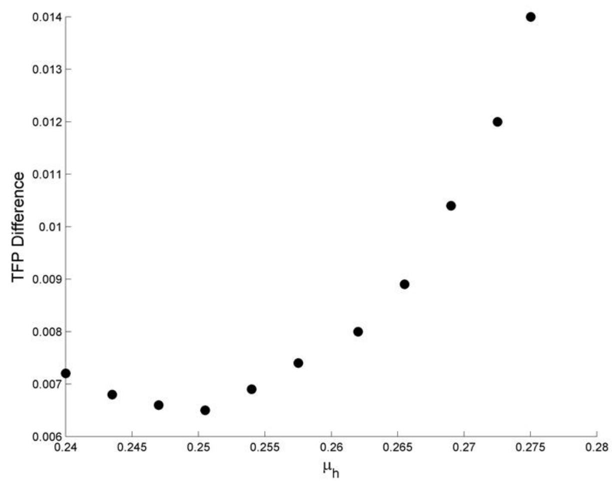


Figure C.14: **Heterogenous Productivity and TFP**

This figure presents the effect of productivity heterogeneity on TFP. On the top it shows TFP in both autarky (red) and financial integration (blue). On the bottom it shows the change of TFP after financial integration

Bibliography

Abreu, D., Pearce, D., and Stacchetti, E. (1990), "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 58, 1041–1063.

Aggarwal, R. and Samwick, A. (1999), "The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation," *Journal of Political Economy*, 107, 65–105.

Ai, H. and Li, R. (2013), "Investment and CEO Compensation under Limited Commitment," Working Paper. University of Minnesota and Purdue University.

Ai, H., Kiku, D., and Li, R. (2012), "A Mechanism Design Model of Firm Dynamics: The Case of Limited Commitment," Working Paper. University of Minnesota, University of Pennsylvania, and Purdue University.

Albuquerque, R. and Hopenhayn, H. (2004), "Optimal lending contracts and firm dynamics," *Review of Economic Studies*, 71, 285–315.

Allen, F., Qian, M., and Qian, J. (2005), "Law, Finance, and Economic Growth in China," *Journal of Financial Economics*, 77, 57–116.

Alvarez, F. and Jermann, U. (2000), "Efficiency, equilibrium, and asset pricing with risk of default," *Econometrica*, 68, 775–797.

Angeletos, G.-M. and Panousi, V. (2011), "Financial integration, entrepreneurial risk and global dynamics," *Journal of Economic Theory*, 146, 863–896.

Atkeson, A. (1991), "International lending with moral hazard and risk of repudiation," *Econometrica*, 59, 1069–1089.

Atkeson, A., Eisfeldt, A., and Weill, P.-O. (2013), "Measuring the Financial Soundness of US Firms 1926-2012," Working paper. University of California-Los Angeles.

Axelson, U. and Baliga, S. (2009), "Liquidity and Manipulation of Executive Compensation Schemes," *Review of Financial Studies*, 22, 3907–3939.

Axelson, U. and Bond, P. (2012), "Wall Street occupations: An equilibrium theory of overpaid jobs," Working paper. London School of Economics and University of Minnesota.

Ayyagari, M., Demirguc-Kunt, A., and Maksimovic, V. (2010), "Formal versus informal finance: Evidence from China," *Review of Financial Studies*, 23, 3048–3097.

Bai, Y. and Zhang, J. (2010), "Solving the Feldstein-Horioka Puzzle with Financial Frictions," *Econometrica*, 78, 603–632.

Bai, Y. and Zhang, J. (2012), "Financial Integration and International Risk Sharing," *Journal of International Economics*, 86, 17–32.

Bayoumi, T., Tong, H., and Wei, S.-J. (2010), "Bonus culture: Competitive pay, screening, and multitasking," Working paper. NBER.

Bebchuk, L. A. and Fried, J. M. (2006), *Pay without performance: The unfulfilled promise of executive compensation*, Harvard University Press.

Bénabou, R. and Tirole, J. (2013), "Bonus culture: Competitive pay, screening, and multitasking," Working paper. Princeton University and Toulouse School of Economics.

Bernanke, B. S. (2005), "The Global Saving Glut and the U.S. Current Account Deficit," Speech at the Sandridge Lecture, Virginia Assoc. Economists, Richmond, VA.

Bernanke, B. S. (2007), "Global Imbalances: Recent Developments and Prospects," Speech at the Bundesbank Lecture, Berlin, Germany.

Bertrand, M. and Mullainathan, S. (2001), "Are CEOs Rewarded for Luck? The Ones Without Principals Are," *Quarterly Journal of Economics*, 116, 901–932.

Biais, B., Mariotti, T., Rochet, J.-C. R., and Villeneuve, S. (2010), "Large risks, limited liability, and dynamic moral hazard," *Econometrica*, 78, 73–118.

Bijlsma, M., Boone, J., and Zwart, G. (2012), "Competition for traders and risk," Working paper, Tilburg University.

Bloom, N. (2009), "The impact of uncertainty shocks," *Econometrica*, 77, 623–685.

Bolton, P., Chen, H., and Wang, N. (2011), "A Unified Theory of Tobin's q, Corporate Investment, Financing, and Risk Management," *Journal of Finance*, 66, 1545–1578.

Bolton, P., Chen, H., and Wang, N. (2013), "Market Timing, Investment, and Risk Management," *Journal of Financial Economics*, 109, 40–62.

Bond, P. and Glode, V. (2013), "Bankers and Regulators," Working paper, University of Washington and University of Pennsylvania.

Brunnermeier, M. and Sannikov, Y. (2012), "A Macroeconomic Model with a Financial Sector," Working paper, Princeton University.

Buera, F., Kaboski, J., and Shin, Y. (2011), "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101, 1964–2002.

Bulow, J. and Rogoff, K. (1989), "Sovereign Debt: Is to Forgive to Forget?" *American Economic Review*, 79, 43–50.

Caballero, R. and Krishnamurthy, A. (2009), "Global Imbalances and Financial Fragility," Working Paper, NBER.

Caballero, R., Farhi, E., and Gourinchas, P.-O. (2008), "An Equilibrium Model of "Global Imbalances" and Low Interest Rates," *American Economic Review*, 98, 358–393.

Campello, M., Giambona, E., Graham, J., and Harvey, C. (2011), "Liquidity Management and Corporate Investment During a Financial Crisis," *Review of Financial Studies*, 24, 1944–1979.

Clementi, G. L. and Hopenhayn, H. (2006), "A theory of financing constraints and firm dynamics," *Quarterly Journal of Economics*, 121, 229–265.

Cooley, T., Marimon, R., and Quadrini, V. (2004), "Aggregate consequences of limited contract enforceability," *Journal of Political Economy*, 112, 817–847.

Core, J. and Guay, W. (2002), "The Other Side of the Trade-off: The Impact of Risk on Executive Compensation: A Revised Comment," Working Paper, University of Pennsylvania.

Cox, D. and Miller, H. (1977), *The theory of stochastic processes*, Chapman and Hall.

DeMarzo, P. and Fishman, M. (2007), "Optimal long-term financial contracting," *Review of Financial Studies*, 20, 2079–2128.

DeMarzo, P. and Sannikov, Y. (2006), "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance*, 61, 2681–2724.

DeMarzo, P., Fishman, M., He, Z., and Wang, N. (2012), "Dynamic Agency and the Q Theory of Investment," *Journal of Finance*, 67, 2295C2340.

Di Tella, S. (2013), "Uncertainty Shocks and Balance Sheet Recessions," Working paper. Stanford University.

Du, J., Lu, Y., and Tao, Z. (2012), "Contracting institutions and vertical integration: Evidence from Chinas manufacturing firms," *Review of Financial Studies*, 40, 89–107.

Eisfeldt, A. and Kuhnen, C. (2013), "CEO Turnover in a Competitive Assignment Framework," *Journal of Financial Economics*, 109, 351–372.

Eisfeldt, A. and Rampini, A. (2006), "Capital Reallocation and Liquidity," *Journal of Monetary Economics*, 53, 369–399.

Ferguson, N. and Schularick, M. (2007), ""Chimerica" and the Global Asset Market Boom," *International Finance*, 10, 215–239.

Fong, Y.-f. and Li, J. (2012), "Relational Contracts, Efficiency Wages, and Employment Dynamics," Working paper. HKUST and Northwestern University.

Frydman, C. and Jenter, D. (2010), "CEO Compensation," *Annual Review of Financial Economics*, 2, 75–102.

Glindro, E., Subhanij, T., Szeto, J., and Zhu, H. (2007), "Are Asia-Pacific Housing Prices Too High For Comfort?" Working Paper, Bank of International Settlements.

Glode, V. and Lowery, R. (2013), "Informed Trading and High Compensation in Finance," Working Paper, University of Pennsylvania and University of Texas-Austin.

Glode, V., Green, R. C., and Lowery, R. (2012), "Financial Expertise as an Arms Race," *Journal of Finance*, 67, 1723–1759.

Gochoco-Bautista, M. S. (2008), "Asset Prices and Monetary Policy: Booms and Fat Tails in East Asia," Working Paper, Bank of International Settlements.

Gourinchas, P.-O. and Jeanne, O. (2013), "Capital Flows to Developing Countries: The Allocation Puzzle," *Review of Economic Studies*, 80, 1484–1515.

Grochulski, B. and Zhang, Y. (2011), "Optimal risk sharing and borrowing constraints in a continuous-time model with limited commitment," *Journal of Economic Theory*, 146, 2356–2388.

Grochulski, B. and Zhang, Y. (2013), "Market-based incentives," Working paper. Federal Reserve Bank of Richmond and University of Texas-Austin.

He, Z. (2009), "Optimal executive compensation when firm size follows geometric brownian motion," *Review of Financial Studies*, 22, 859–892.

He, Z. and Krishnamurthy, A. (2013a), "Intermediary Asset Pricing," *American Economic Review*, 103, 732–770.

He, Z. and Krishnamurthy, A. (2013b), "A Macroeconomic Framework for Quantifying Systemic Risk," Working Paper, University of Chicago and Northwestern University.

He, Z. and Xiong, W. (2012), "Rollover Risk and Credit Risk," *Journal of Finance*, 67, 391–430.

Hoffmann, F. and Pfeil, S. (2010), "Reward for Luck in a Dynamic Agency Model," *Review of Financial Studies*, 23, 3329–3345.

Holmstrom, B. (1982), "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

Holmstrom, B. and Milgrom, P. (1987), "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica*, 55, 303–328.

Hsieh, C.-T. and Klenow, P. (2009), "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

Igan, D. and Jin, H. (2010), "Asian Real Estate Markets: On Bubble Alert?" Global Financial Stability Report, International Monetary Fund.

IMF (2007), "Global Capital Flows: Defying Gravity, Finance and Developement," .

Jenter, D. and Kanaan, F. (2010), "CEO Turnover and Relative Performance Evaluation," Working paper. Stanford University.

Kaplan, S. (2012), "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts and Challenges," Working paper. University of Chicago.

Kaplan, S. and Minton, B. (2012), "How Has CEO Turnover Changed," *International Review of Finance*, 12, 57–87.

Kaplan, S. and Rauh, J. (2010), "Wall Street and Main Street: What contributes to the rise in the highest incomes," *Review of Financial Studies*, 23, 1004–1050.

Kehoe, P. and Perri, F. (2002), "International business cycles with endogenous incomplete markets," *Econometrica*, 70, 907–928.

Kehoe, T. and Levine, D. (1993), "Debt-constrained asset markets," *Review of Economic Studies*, 60, 865–888.

Kiyotaki, N. and Moore, J. (1997), "Credit Cycles," *Journal of Political Economy*, 105, 211–248.

Kramer, C. (2006), "Asia's Investment Puzzle," Working Paper, International Monetary Fund.

Krishnamurthy, A. (2003), "Collateral Constraints and the Amplification Mechanism," *Journal of Economic Theory*, 111, 277–292.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (1997), "Legal Determinants of External Finance," *Journal of Finance*, 52, 1131–1150.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (1998), "Law and Finance," *Journal of Political Economy*, 106, 1113–1155.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (2000), "Investor Protection and Corporate Governance," *Journal of Financial Economics*, 58, 3–27.

Lambert, R. and Larcker, D. (1987), "An analysis of the use of accounting and market measures of performance in executive compensation contracts," *Journal of Accounting Research*, 25, 85–125.

Leland, H. E. (1994), "Corporate Debt Value, Bond Covenants, and Optimal Capital Structure," *Journal of Finance*, 49, 1213–1252.

Leland, H. E. and Toft, K. B. (1996), "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads," *Journal of Finance*, 51, 987–1019.

Levin, J. (2003), "Relational Incentive Contracts," *American Economic Review*, 93, 835–857.

Li, R. (2012), "Optimal Contract Design over the Business Cycle," Working paper, Purdue University.

Lorenzoni, G. (2008), "Inefficient Credit Booms," *Review of Economic Studies*, 75, 809–833.

Lu, Y. and Tao, Z. (2009), "Contract enforcement and family control of business: Evidence from China," *Journal of Comparative Economics*, 37, 597–609.

Ma, G. and Yi, W. (2010), "China's High Saving Rate: Myth and Reality," Working Paper, Bank for International Settlements.

Maggiori, M. (2012), "Financial Intermediation, International Risk Sharing, and Reserve Currencies," Working Paper, New York University.

Mendoza, E., Quadrini, V., and Rios-Rull, J.-V. (2009), "Financial Integration, Financial Development, and Global Imbalances," *Journal of Political Economy*, 117, 371–416.

Miao, J. and Zhang, Y. (2013), "A Duality Approach to Continuous-Time Contracting Problems with Limited Commitment," Working paper, Boston University and University of Texas-Austin.

Midrigan, V. and Xu, D. Y. (2011), "Finance and Misallocation: Evidence from Plant-level Data," Working paper, NBER.

Moll, B. (2013), "Productivity Losses from Financial Frictions: Can Self-financing Undo Capital Misallocation?" Working paper, Princeton University.

Opp, M. and Zhu, J. (2013), "Impatience vs. Incentives," Working paper, University of California-Berkeley and University of Pennsylvania.

Oyer, P. (2008), "The making of an investment banker: Stock market shocks, career choice, and lifetime income," *Journal of Finance*, 63, 2601–2628.

Peters, F. and Wagner, A. (2013), "The Executive Turnover Risk Premium," *Journal of Finance*, forthcoming.

Philippon, T. and Reshef, A. (2012), "Wages and Human Capital in the US Finance Industry: 1909C2006," *Quarterly Journal of Economics*, 127, 1551–1609.

Piskorski, T. and Tchistyi, A. (2010), "Optimal mortgage design," *Review of Financial Studies*, 23, 3098–3140.

Prendergast, C. (2002), "The Tenuous Trade-off between Risk and Incentives," *Journal of Political Economy*, 110, 1071–1102.

Rajan, R. G. (2006a), "Investment Restraint, the Liquidity Glut, and Global Imbalances," Remarks at the Conference on Global Imbalances, Bali.

Rajan, R. G. (2006b), "Is There a Global Shortage of Fixed Assets?" Remarks at the G-30 Meetings in New York City.

Rampini, A. and Viswanathan, S. (2013a), "Collateral and capital structure," *Journal of Financial Economics*, 109, 466–492.

Rampini, A. and Viswanathan, S. (2013b), "Collateral, Risk Management, and the Distribution of Debt," *Journal of Finance*, 65, 2293–2322.

Ray, D. (2002), "The Time Structure of Self-Enforcing Agreements," *Econometrica*, 70, 547–582.

Sannikov, Y. (2008), "A continuous-time version of the principal-agent problem," *Review of Economic Studies*, 75, 957–984.

Schwarz, W. (1992), "The Wiener Process between a Reflecting and an Absorbing Barrier," *Journal of Applied Probability*, 65, 597–604.

Song, Z., Storesletten, K., and Zilibotti, F. (2012), "Growing Like China," *American Economic Review*, 101, 196–233.

Spear, S. and Srivastava, S. (1987), "On repeated moral hazard with discounting," *Review of Economic Studies*, 54, 599–617.

Thanassoulis, J. (2012), "The case for intervening in bankers' pay," *Journal of Finance*, 67, 849–895.

Thomas, J. and Worrall, T. (1988), "Self-Enforcing Wage Contracts," *Review of Economic Studies*, 61, 541–554.

Thomas, J. and Worrall, T. (1994), "Foreign direct investment and the risk of expropriation," *Review of Economic Studies*, 61, 81–108.

Ward, A. and Glynn, P. (2003), "Properties of the Reflected Ornstein–Uhlenbeck Process," *Queueing Systems*, 44, 109–123.

Wu, J., Gyourko, J., and Deng, Y. (2012), "Evaluating conditions in major Chinese housing markets," *Regional Science and Urban Economics*, 42, 531–543.

Yang, D. T., Zhang, J., and Zhou, S. (2010), "Why are Saving Rates so High in China," Working Paper, Hong Kong Institute for Monetary Research.

Zhang, L. and Du, Z. (2010), "On the Reflected Geometric Brownian Motion with Two Barriers," *Intelligent Information Management*, 2, 295–298.

Zhu, J. Y. (2013), "Optimal Contracts with Shirking," *Review of Economic Studies*, 80, 812–839.

Biography

Felix Zhiyu Feng (born on September 28th, 1984, in Chengdu, China) is a Doctor of Philosophy candidate in economics at Duke University. He holds a B.A in economics and finance from Peking University, China in 2007, and a M.A. in economics from Duke University in 2008. His Ph.D. work focuses on theoretical corporate finance, especially in the areas of corporate governance and investments. He received the Tsang Hin-Chi Undergraduate Scholarship in 2004, the Distinguished Undergraduate Thesis Award in 2007, the Fooster-Comes Research Fellowship in 2013, and the Duke University Graduate School Fellowship from 2009 to 2014. He will join the Department of Economics at the University of Notre Dame as an assistant professor in August, 2014.