MIT Sloan School of Management

MIT Sloan School Working Paper 5358-18

Customers and Investors: A Framework for Understanding the Evolution of Financial Institutions

Robert C. Merton and Richard T. Thakor

Original June 1, 2015 MIT Sloan WP 5137-15
This version March 17, 2018

Customers and Investors: A Framework for Understanding the Evolution of Financial Institutions*

Robert C. Merton† and Richard T. Thakor‡

Forthcoming, *Journal of Financial Intermediation*

Abstract

Financial institutions are financed by both investors and customers. Investors expect an appropriate risk-adjusted return for providing financing and risk bearing. Customers, in contrast, provide financing in exchange for specific services, and want the service fulfillment to be free of the intermediary's credit risk. We develop a framework that defines the roles of customers and investors in intermediaries, and use it to build an economic theory that has the following main findings. First, with positive net social surplus in the intermediary-customer relationship, the efficient (first best) contract completely insulates the customer from the intermediary's credit risk, thereby exposing the customer only to the risk inherent in the contract terms. Second, when intermediaries face financing frictions, the second-best contract may expose the customer to some intermediary credit risk, generating "customer contract fulfillment" costs. Third, the efficiency loss associated with these costs in the second best rationalizes government guarantees like deposit insurance even when there is no threat of bank runs. We further discuss the implications of this customer-investor nexus for numerous issues related to the design of contracts between financial intermediaries and their customers, the sharing of risks between them, ex ante efficient institutional design, regulatory practices, and the evolving boundaries between banks and financial markets.

Key words: Customers, investors, credit risk, financial intermediaries, real-world financial contracts, information-insensitivity

JEL Classification Numbers: D81, D83, G20, G21, G22, G23, G24, G28, H81

---

* Original Draft: June, 2015. For their helpful comments, we thank Franklin Allen (discussant), Doug Diamond, Zhiguo He (discussant), Stavros Zenios, an anonymous referee, the JFI editors, and participants at the Federal Reserve Bank of Cleveland and the Office of Financial Research Financial Stability Conference and the WFA-CFAR/JFI Conference on "The Post-Crisis Evolution of Banks and Financial Markets". We alone are responsible for remaining errors, if any.

† MIT Sloan School of Management and NBER

‡ University of Minnesota, Carlson School of Management

1. Introduction

Many financial intermediaries provide services whose effective delivery depends on the creditworthiness of the provider. Merton (1989, 1993, 1997) defines these as "credit-sensitive" financial services, and points out that the intermediary's credit standing can generate externalities for the different business activities of the intermediary because of these services, even when the business activities are not directly interconnected through common customers or other means.[1] An example is an investment bank that considers participating in a bridge loan to start a merchant banking business, and in doing so risks having institutional customers flee its over-the-counter derivatives business (e.g. long-dated swap contracts) because of concerns about the bank's ability to fulfill its contractual obligations on its derivative products were it to suffer losses on its bridge loans (see Merton (1997)). In financial intermediation theories, the *raison d'etre* of a financial institution is to serve its customers (policyholders in the case of an insurance company, for example), so the potential sensitivity of the perceived value to customers of the intermediary's services to its own credit risk has important implications.

An example of a credit-sensitive financial service is the vector of services banks provide their depositors. The implication of depositor exposure to the bank's credit risk that has been explored in one strand of the literature has to do with the desirability of riskfree deposits. This literature has suggested that uninformed and risk-averse depositors demand riskfree deposits because this either eliminates their disadvantage in trading with informed agents (e.g. Gorton and Pennacchi (1990)) or improves risk sharing (e.g. Dang, Gorton, Holmstrom, and Ordonez (2017)).

In this paper, we argue that the bank deposit contract is but one example of a much broader set of contracts between financial intermediaries and their customers in which customers would prefer

[1] See also Merton (1990, 1992a). These papers are part of the "functional perspective" of financial intermediation.

to be insulated from the credit risks of the intermediaries they are contracting with, even when they are not risk averse. The basic idea in our analysis is that the intermediary-customer relationship produces non-tradeable consumer surplus whose expected value declines when the intermediary's probability of bankruptcy/liquidation increases, signifying a diminished ability to serve the customer. Moreover, this counterparty-risk problem cannot be resolved by having the customer purchase insurance against intermediary failure. In fleshing out this idea in our theory, our contribution is that we provide a broader "functional perspective" on the relationship between a financial intermediary and its customers, and how this affects contract design, institutions, and regulation. This exercise builds on Merton's (1989, 1993, 1995, 1997) insights, but goes beyond them in providing a formal analysis of efficient contractual arrangements as well as deviations from efficiency due to contracting frictions. Moreover, we juxtapose these insights with the growing literature on the role of banks in "manufacturing safe debt" via deposit contracts that insulate depositors from the bank's risks. This enables us to explain existing contractual arrangements and important recent regulatory practices.

Specifically, the questions we address in this paper are: what are the implications of this customer-investor nexus for how the financial intermediaries structure efficient (first-best) contracts with their customers? That is, *why* do customers not wish to be exposed to intermediary credit risk? When financing frictions impede the adoption of efficient contracts, how does this perspective illuminate the microfoundations of observed (second-best) contracts between intermediaries and their customers? What implications does this have for certain institutional arrangements and regulations and the evolving boundary between banks and financial markets?

In addressing these questions, the starting point of our analysis is that financial institutions differ from non-financial firms in at least two noteworthy respects. First, a financial institution's

investors purchase claims that look similar to what its customers purchase, e.g. subordinated debtholders (investors) and depositors (customers) who both have debt claims on the bank. By contrast, customers in a non-financial firm like IBM purchase products that are transparently different from the claims of its investors. Second, in financial institutions both investors and customers provide financing to the intermediary.[2] Investors, like shareholders and bondholders, provide financing and risk bearing since the values of their claims are linked to the intermediary's outcomes. Customers, in contrast, expect services in exchange for the financing they provide, but prefer not to bear intermediary-specific credit risk, i.e., they want the intermediary's service provision to not depend on the fortunes of the service provider.[3]

We distinguish between two types of customers in financial intermediaries: "credit-sensitive" customers and "other" customers. "Credit-sensitive" customers provide financing *to* the intermediary in exchange for future services; this financing is a liability of the intermediary. The utility customers derive from the intermediary's services is diminished by an increase in the credit risk of the intermediary. "Other" customers are those who receive financing *from* the intermediary, such as bank borrowers. They appear on the asset side of the intermediary's balance sheet, and are *not* credit-sensitive since they are obliged to repay the intermediary in the future. Our focus is on "credit-sensitive" customers (we refer to them as just "customers" henceforth). We show that the additional expected return required to induce them to bear the credit risk of the intermediary exceeds that required to induce the investors to bear it. Thus, a financial intermediary that imposes credit risk on its customers will not be able to compete effectively against one that does not. For

[2] In non-financial firms, *suppliers* provide the firm with trade credit, which is short-term financing in the form of payables. However, customers end up being *consumers* of finance rather than providers of it. In contrast, in the case of commercial banks, deposits represent customer-financing and make up typically 70%-80% of the bank's total financing.

[3] For example, a life insurance company's policyholders are customers who provide cash premiums to finance the company's assets, but also create liabilities for the insurance company. Similarly, depositors in a bank provide (debt) financing for the bank, but they are also consumers of a variety of safekeeping, liquidity and transaction services.

example, for a whole-life policyholder in a life insurance company to be indifferent to a lowering of the likelihood that the policy will pay off in the event of death, the insurance company will have to increase the expected return on the customer's investment more than it would have to if it imposed this risk on its investors instead. This sheds light on some survey evidence. Wakker, Thaler, and Tversky (1997) report that respondents in their surveys said they would pay 20% less for an insurance policy if the probability of default by the insurance company rose from 0% to 1%. Wakker, Thaler, and Tversky (1997) argue that this is hard to reconcile with standard expected utility theory. We provide a rational explanation for such behavior.

The key here is not the identity of the economic agent, but the *role* played by that agent, i.e., whether the agent is an investor or a customer who also provides financing. In some instances, the agent may play multiple roles, and may therefore have different expectations of the institution in different roles, e.g., a policyholder in an insurance company is a customer but may also hold the company's stock as an investor. This clarifies that the focus of our analysis is *not* on the primitives associated with economic agents—such as their preferences, beliefs, or wealth endowments—but rather what they view as the optimal contract between them and the intermediary in a given role.[4] Another key is that failure of the intermediary may lead to liquidation and the inability to provide full service to the customer (e.g. see Allen and Gale (2009)).[5] If bankruptcy is merely reorganization that does not affect the customer, there would be no efficiency loss. However, this is typically not the case.

Corresponding to the questions listed earlier, our main results can be summarized as follows. First, we analyze the efficient (first-best) contract between the intermediary and the customer and

[4] For example, an individual will be a customer of a bank in which he/she has a retail deposit account and an investor with respect to the purchase of stocks of publicly-traded firms.
[5] As in the case of the Lehman Brothers bankruptcy in 2008.

show that as long as the contract creates positive net social surplus, it completely insulates the customer from the credit risk of the intermediary. Consequently, the customer is exposed only to the risk stipulated in the contract terms, and not the credit risk of the intermediary itself. We show that exposing the customer to the intermediary's credit risk is akin to affixing to the contract a lottery that has negative social value, and that because of this all of the intermediary's credit risk is borne by its investors in the efficient contract. We further show that asking the customer to diversify exposure to the intermediary's credit risk by purchasing contracts from a large number of intermediaries is inefficient relative to the intermediary's investors bearing this risk. A key element of the argument is that the customer operates in an inherently incomplete market while purchasing a contract from a financial intermediary. However, our argument does not rely on any lack of sophistication on the part of customers, risk aversion, or constrained access to information—the customers in our analysis are not simply "widows and orphans" or uninformed/unsophisticated investors. A customer could be an institution such as the World Bank or a large pension fund. We also show that a financial contracting solution like having the customer purchase a guarantee that compensates the customer for the loss in utility due to the intermediary's failure does *not* affect the welfare loss due to intermediary credit risk.

Second, we analyze the second-best contract, which is constrained-efficient in the sense that the intermediary may face costly financing frictions that obstruct its ability to completely insulate the customer from the intermediary's credit risk. In this case, there is a tradeoff between the loss of efficiency (relative to the first-best) from exposing the customer to the intermediary's credit risk on the one hand, and the cost of insulating the customer from this credit risk on the other hand. The second-best contract may thus expose the customer to some of the credit risk of the

intermediary, absent government intervention.[6] Indeed, some government intervention may be rationalized by the goal of reducing these costs. The loss of efficiency in the second best is referred to as "customer contract fulfillment" (CCF) costs.

Third, we discuss how our analysis explains a variety of observed real-world contracts, institutions, and regulatory practices. The contracts that are rationalized by the framework developed in this paper are: insured bank deposits, mutual funds, insurance contracts, and repos in shadow banking.[7] In all of these examples, we describe who the customers are, why they would care about intermediary credit risk, and why government intervention may sometimes be necessary. An institution we analyze is a futures exchange, and we explain how an exchange insulates customers (the holders of contracts) from counterparty risk and why this enhances welfare. Our analysis offers insights into some regulatory practices in banking, specifically the Dodd-Frank Act enacted in 2010 in response to the 2008-2009 global financial crisis. The element of this regulation that we provide economic foundation for is the requirement for swaps to be traded through clearing houses and exchanges, and we explain how this helps to protect customers from being exposed to intermediary credit risk. We also explore how our framework provides a perspective on the role of the government in reducing CCF costs, thereby illuminating policies like "too big to fail".

Finally, we also analyze how the boundary between banks and financial markets becomes blurred as banks choose more market-based activities. We use our framework to develop a simple model in which banks choose the extent to which they want to integrate themselves with financial markets and show that the sensitivity of the bank's depository customers to the bank's credit risk leads the bank to curtail the extent of integration, whereas higher regulatory costs may push banks

[6] This explains why customers are sometimes willing to deal with institutions that do not have a AAA credit rating.
[7] For mutual funds, the exception is if the mutual fund is providing liquidity services for cash.

in the opposite direction.

The rest of this paper is structured as follows. Section 2 briefly reviews the related literature. In Section 3, we present the basic framework of a financial intermediary with investors and customers to develop a theory that enables a characterization of the first-best contract. We introduce financing frictions for the intermediary in Section 4 to explain why such separation between the contract and the credit risk of the intermediary can be less than perfect in the second-best contract. We show how the optimal degree of exposure of the customer to the credit risk of the intermediary in the second-best contract is determined and how this generates CCF costs. Section 5 turns to a discussion of how the analysis illuminates observed contracts, institutions, and regulations. Section 6 examines how the theory sheds light on the evolving boundaries between banks and financial markets. Section 7 provides concluding remarks.

2. Related Literature and Contribution

Broadly speaking, there are four related strands of the literature: the literature on the *demand* for information-insensitive (and safe) debt contracts like banks deposits, the literature on the existence of financial intermediaries in which safe debt is a consequence of the intermediary being infinitely large in equilibrium, the security design literature that explains why firms may wish to *supply* safe debt, and the functional perspective of financial intermediaries. We discuss these strands here and explain what this paper adds at the margin.

Consider the first strand. Gorton and Pennacchi (1990) first proposed that agents who lack the skills to efficiently acquire and process information would prefer to invest in instruments like bank deposits that are informationally insensitive so as not to be disadvantaged in trading with informed agents. Since then, others have rationalized debt contracts that are informationally-insensitive to

provide optimal risk sharing. For example, Dang, Gorton, Holmstrom, and Ordonez (2017) rely on the Hirshleifer (1971) notion that information may sometimes not be released because its release can distort risk sharing.[8] In their model, there are two generations of (globally) risk-averse depositors. The early generation of depositors want to sell their claims to the late generation if hit by a liquidity shock, but at a non-random price, which means they do not want the late depositors to produce bank-asset-value information that makes their exit price information-contingent. The bank will oblige by withholding information and investing in opaque assets that discourage information production; this makes deposits information-insensitive.

A different perspective on why bank deposits are (optimally) riskless, at least asymptotically, is provided by the second strand of the literature that provides the information-based microfoundations for financial intermediary existence. In both Diamond (1984) and Ramakrishnan and Thakor (1984), the intermediary efficiently diversifies away the idiosyncratic risks of individual loans/projects, so that even if an individual loan that is monitored/screened by the bank remains (partially) opaque, the bank itself becomes riskless as it grows to its efficient size. The optimality of such an intermediary does not depend on depositor risk aversion, however.[9] Not relying on risk aversion to explain the demand for safe debt is also consistent with the stylized fact that investors are willing to pay a "premium" for riskless debt by accepting a lower yield than implied by risk aversion (e.g. Krishnamurthy and Vissing-Jorgensen (2012)).[10]

These papers focus on the demand for safe debt. A third strand of the literature provides a

[8] See also Holmstrom (2015).

[9] These papers reach essentially the same conclusion of riskfree deposits as Dang, Holmstrom, Gorton, and Ordonez (2017), but reverse the causality in the argument—the bank is not opaque because it wants to appear informationally-insensitive to its depositors, but rather it diversifies away the idiosyncratic risk associated with each individually-opaque asset it monitors/screens in order to reduce contracting costs and thus asymptotically eliminate risk for its depositors, so that its overall asset portfolio is indeed transparently riskfree to its depositors.

[10] A number of recent papers emphasize the special role of banks in liquidity creation and assign a liquidity premium to safe debt, e.g. DeAngelo and Stulz (2015), Hanson, Shleifer, Stein, and Vishny (2015), and Hart and Zingales (2014).

supply-side perspective and appears in the security design literature in which firms engage in tranching their total cash flows to produce multiple claims, some that are less information-insensitive than the total cash flows and others that are more information-sensitive. For example, Boot and Thakor (1993) develop a theory of security design in which riskless debt and information-sensitive equity emerge as optimal contracts for an issuer maximizing expected revenue from issuing securities. Their model also explains asset pooling, securitization, and tranching. DeMarzo and Duffie (1999) develop a model in which an issuer raises capital by securitizing part of its assets. The issuer's private information at the time of security issuance causes illiquidity in the security. The paper characterizes conditions under which standard debt is an optimal security.[11] Thus, rather than focusing on customers' needs, this literature focuses on how safe securities created via tranching serve security issuers.

The fourth strand is the literature on the functional perspective in finance (for example, Merton (1990, 1993, 1995), and Merton and Bodie (1995, 2005); see Campbell and Wilson (2014) for a review). Consistent with this literature, our focus is on the functions that financial intermediaries serve in meeting customer needs, and we show that these customers are best served when insulated from the intermediary's credit risk. Thus, while we examine specific contracts and institutions as applications of our framework, these serve mainly as examples of the *functions* we seek to highlight in the customer-intermediary interaction.

Our theory differs from the first three strands described above in a number of significant ways. First, we sharply distinguish between customers and investors in financial institutions, and show that *only* the customers should be protected from the intermediary's fortunes in an efficient

[11] See also DeMarzo (2005). Fulghieri and Lukin (2001) examine optimal security design and show that the Myers and Majluf (1984) pecking order aversion of firms to equity need not hold when outside investors can produce information about the firm and the equilibrium degree of information asymmetry is endogenous. That is, they provide an information-based rationale for equity, rather than safe debt.

contract. Second, in our framework, it is not only bank deposits that should be optimally insulated from bank credit risk—and hence made insensitive to bank-specific information—but *all* efficient contracts between the financial intermediary and its *customers*. This includes a far bigger set of contracts and institutions than bank deposits. For example, insurance contracts, repos, and futures exchanges are also included. Third, in our framework, the efficient claim of the customer need not be riskless—it can be risky, but the risk must be confined to the promised state-contingent payoffs of the contract itself and *cannot* include the credit risk of the intermediary. Thus, we are not just talking about "safe" debt. Fourth, we address the important question of *why* all of the credit risk and the affiliated informational risk should be borne by the intermediary's investors in the first-best case, and not by its customers. This enables us to shed new light on issues like the need for deposit insurance even in the absence of the threat of contagious bank runs and the Dodd-Frank Act. Fifth, our main finding that the value of the customer's claim must be independent of the credit risk of the intermediary in the first best does not depend on customer risk aversion or on information acquisition by customers being prohibitively costly or inimical to stability. Rather, our approach suggests that in well-functioning markets, customers do not have a *need* for their contracts to be opaque, since their contracts should be optimally structured to insulate them from the risks of the service-providing intermediaries. While opaqueness may benefit producers (e.g. banks), our analysis suggests that it need not benefit customers who will be indifferent to opaqueness as long as their claims do not depend on the fortunes of the intermediary. Therefore, in high-quality debt markets, it need not be the case that transparency causes dysfunction or that opaqueness is necessary. Finally, our analysis of the second best highlights the potential channels through which financing frictions can diminish efficiency in the customer-intermediary contract by imposing intermediary-specific credit risk on the customer, and the resulting CCF costs.

The marginal contributions of our paper relative to the fourth strand of the literature—the functional perspective—can be described as follows. First, Merton (1989, 1993, 1997) focuses on the efficient (first-best) contract. We formally analyze this contract and characterize its properties, and establish a new result that having the customer purchase a guarantee to be compensated for the loss in utility from intermediary default does not reduce the welfare loss from the customer's exposure to intermediary credit risk. Second, we highlight the financial frictions that may result in the contract not always being encountered in practice, and we describe the resulting loss in efficiency as a CCF cost in the second-best contract. The characterization of the CCF cost is novel to this paper. Finally, we explain how our analysis can shed light on numerous contracts, institutions, and regulatory practices. For example, it rationalizes federal deposit insurance even if there is no threat of bank runs, and the blurring of boundaries between banks and markets.

3. Financial Intermediaries and Customers

3.1 Analytic Setting: Efficient Customer Contracts (First-Best)

We now introduce a simple analytic example to define and discuss the key concepts concretely. Let V be the value of the service that an intermediary provides to its customer. It is the monetary equivalent of the expected utility (or the certainty-equivalent of the expected utility) that the customer gets at $t = 0$ from the intermediary's services, and can have many components, as we discuss below. Thus, if the customer is a depositor, then V could represent the monetary equivalent of the value the depositor attaches to having access to a liquid claim at a moment's notice. For a policyholder in an insurance company, V could represent the value of the utility the individual derives from being able to insure against an accident or a catastrophic event like death. In all of these cases, the contract calls for the customer to provide a set of payments f_t to the intermediary

at various dates $t \in [0,T]$, where $[0,T]$ is the period over which the contract exists, in exchange for a vector of services that may include future monetary payments.

More specifically, from the perspective of the customer, *V* includes two components. The first component is V_m, which is the monetary equivalent of the utility that the customer derives based on the net *monetary flows* between the customer and the intermediary—i.e., the money f_t flows from the customer to the intermediary, and the (possibly state-contingent) money *F* that is paid by the intermediary to the customer as part of the service provided by the intermediary. In a bank, f_t is the customer's deposit in the bank at date *t*, and *F* the amount of deposits (plus interest) withdrawn by the depositor.[12] In an insurance context, f_t would represent the vector of insurance premia paid to the insurance company and *F* the payment made by the insurance company in the event of an accident or death. While *F* may be deterministic, it can also be stochastic. The second component is V_s, which is the monetary equivalent of the utility the customer derives from the *services* provided by the intermediary. As an example, if the customer is a bank depositor, then V_m would be the monetary equivalent of the depositor's utility from receiving interest on the deposit (the difference between what the bank returns to the depositor and what was deposited in the bank), whereas V_s would include the monetary equivalent of the utility associated with check-writing privileges, access to liquidity (including states in which such liquidity may be unavailable elsewhere), safe-keeping services, cash management advice, etc. Put together, the two components sum up to *V*, so $V_m + V_s = V$.

Now define $\overline{V}$ to be the monetary equivalent of the reservation utility of the customer—it will capture the opportunity cost for the customer to use the financial intermediary rather than purchase

[12] Put another way, V_m is the monetary equivalent value of the expected utility from $F - \text{PV}(\sum_t f_t)$, which represents the present value of all monetary flows. There need not be only one withdrawal. With multiple withdrawals, *F* would be the present value of all withdrawals.

the service through, say, another intermediary or even the financial market.[13] Satisfaction of the customer's participation constraint requires

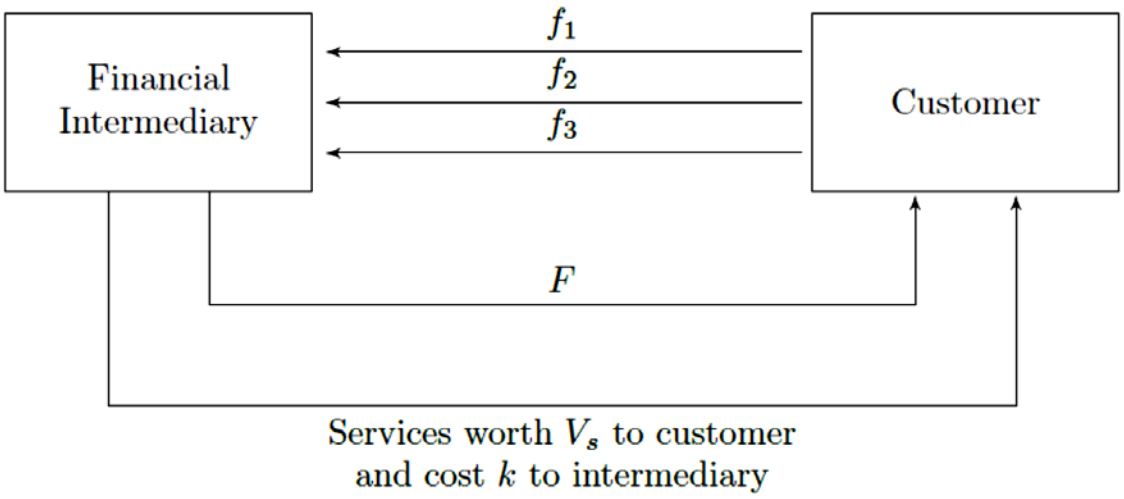
$$V_m + V_s = V \geq \overline{V} \tag{1}$$

Let $k > 0$ be the cost to the intermediary of providing the vector of services that the customer values, and V_m^I the monetary value at $t = 0$ of the intermediary's services. *Figure 1* below describes the relationship between the intermediary and the customer in terms of the values and costs of the financing provided by the intermediary and the value of its services. The figure shows that in the kinds of contracts we are interested in, the customer first provides financing to the intermediary (f_1, f_2, …, etc.) and then the intermediary provides a financial payoff (F) and services (V_s) to the customer at a future date.

[13] Merton (1989) suggests one reason why an intermediary may be able to improve upon the market in providing service to the customer, specifically by providing *customized* derivatives securities that generate a payoff stream that replicates the customer's desired payoff stream emerging from an intertemporal portfolio optimization. Because the intermediary can aggregate derivatives contracts and then hedge risk in the market, the arrangement is more efficient than the individual customer trading directly in the market.

Figure 1: Values and Costs of Financing and Services

This figure illustrates the flows of values and costs between the financial intermediary and the customer. The f_i arrows represent the payments made by the customer to the intermediary. The F arrow represents the financial payoff made by the intermediary to the customer at a future date. The lower arrow represents the services provided by the intermediary to the customer, V_s, at a cost of k to the intermediary.



We assume that

$$V_m^I - k > 0 \tag{2}$$

Taken together, (1) and (2) imply that intermediation creates a positive net economic surplus. This net surplus (in dollars) is

$$V - \overline{V} + V_m^I - k > 0 \tag{3}$$

We can attribute this surplus to the specialization-related skills that provide the economic rationale for the existence of the financial intermediary. Let the duration of the contract between the intermediary and the customer be over the time period $[0,T]$.

For expositional simplicity, suppose the contract is entered into at $t = 0$, at which date the customer provides financing, and then the contract is fulfilled at a single date $t = T$, at which time

the intermediary provides all of the services the customer values at V. Let $p \in [0,1]$ be the probability that the intermediary will be solvent at $t = T$, and only if it is solvent can the services valued by the customer be provided. Thus, $1 - p$, the complement of this probability, represents the idiosyncratic credit risk of the intermediary that the contract is exposed to. The value of the contract to the customer now becomes pV, and the participation constraint now becomes $pV \geq \overline{V}$. Thus, the customer's net expected economic surplus relative to its other options is $pV - \overline{V}$. This net surplus is $V - \overline{V}$ if there is no credit risk, which means that the expected loss of net economic surplus due to the intermediary's credit risk is $[1-p]V$. The total expected value (to both the intermediary and the customer) due to the contract is $pV + V_m^I$, and the total net expected economic surplus considering the intermediary's cost of service provision k and the customer's alternative to the contract is $pV + V_m^I - [\overline{V} + k]$.[14] Absent intermediary credit risk, the net economic surplus is $V + V_m^I - [\overline{V} + k]$. This means that the expected loss of net economic surplus due to the credit risk of the intermediary is $[1-p]V$, which is increasing in the intermediary's credit risk, $[1 - p]$. We call this a "customer contract fulfillment" (CCF) cost. The efficient contract drives this cost down to zero.

As mentioned in the Introduction, we are assuming that when the intermediary is bankrupt (or in financial distress), there is a real consequence in terms of impaired ability to provide liquidity services to depositors, i.e., bankruptcy is not just a frictionless reorganization that leaves depositors unaffected. This is similar to Bernanke (1983), who stresses that the bankruptcy of a bank can destroy loan capabilities.[15] Whereas Bernanke (1983) focused on the borrower side of the effect

[14] V_m^I is not multiplied by p in these expressions because all financing is provided by the customer up front at $t = 0$. Thus, insolvency on the part of the intermediary at a later date will not reduce the value to the intermediary of obtaining financing from the customer.

[15] In explaining why depressed output takes so long to rebound after financial crises, he states: "The basic premise is that, because markets for financial claims are incomplete, intermediation between some classes of borrowers and

of bank failure, we focus on the customer side.

We can now characterize how the total economic value of the contract surplus (to the customer and the intermediary) and the customer's share of this total contract value behave as functions of the intermediary's credit risk, $1 - p$. These relationships are depicted graphically in *Figure 2*.

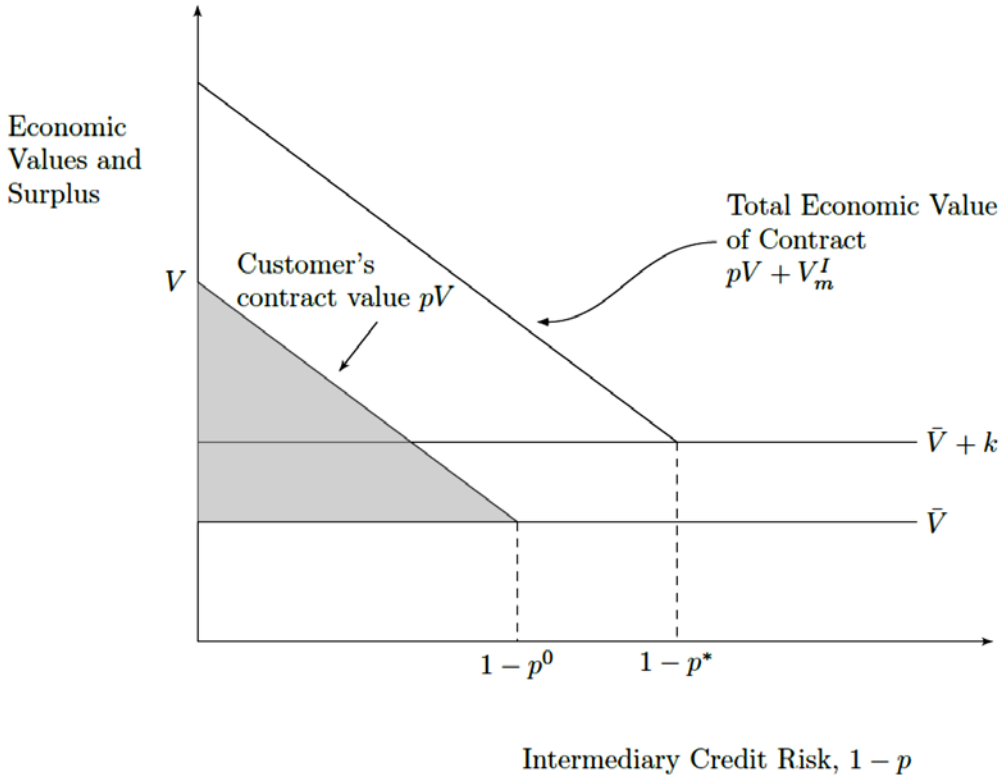
Theorem 1: *The rate at which the total expected net economic surplus,* $ES_{total} = pV + V_m^I - [\overline{V} + k]$, *declines with respect to financial intermediary credit risk is the same as the rate at which the customer's net expected economic surplus,* $ES_c = pV - \overline{V}$, *declines with intermediary credit risk, and this rate is increasing in V. The intermediary credit risk,* $1 - p^0$, *at which* ES_c *becomes zero is less than the credit risk,* $1 - p^*$, *at which* ES_{total} *becomes zero. Moreover,* p^* *is decreasing in* $V_m^I - k$, *the spread between the monetary value of the intermediary's service and the cost of providing that service.*

This result implies that the larger the value of the service provided by the intermediary, the faster is the rate of decline of the economic surplus from the customer relationship due to an increase in intermediary credit risk, i.e., more valuable relationships are more sensitive to intermediary credit risk. The intuition for why the value of the customer's net expected economic surplus becomes zero at a lower level of intermediary credit risk than the level at which total expected net economic surplus becomes zero is that there is also a net producer surplus, $V_m^I - k$, for the intermediary (see *Figure 2* below). Since p^* is decreasing in this surplus, the larger this surplus, the bigger is the spread between $1 - p^*$ and $1 - p^0$. In *Figure 2* below, we show how total surplus and customer surplus decline as the intermediary's insolvency risk $(1 - p)$ increases.

lenders requires non-trivial market-making and information-gathering services. The disruption of 1930-33 […] reduced the effectiveness of the financial sector in performing these services."

Figure 2: The Effect of Intermediary Credit Risk on Contract Value and Expected Net Economic Surplus

This figure shows how intermediary credit risk affects economic surplus. The horizontal axis signifies intermediary credit risk, represented by $1 - p$. The vertical axis signifies expected economic surplus. The customer's expected contract value, pV, is given by the lower decreasing line. The lower horizontal line at $\overline{V}$ represents the reservation utility of the customer, and therefore the shaded region signifies the customer's economic surplus. The higher decreasing line gives the total expected economic value of the contract, accounting for the intermediary's value, $pV + V_m^I$. The lower horizontal line at $\overline{V} + k$ represents the combined customer reservation utility and cost to intermediary, and therefore the distance between this line and the total expected economic value of the contract gives the net economic surplus.



Two points are worth noting. First, if the intermediary exposes the contract to its own credit risk, the customer cannot recover the entire loss of surplus by hedging this risk—say by buying a put option on the intermediary. The reason is that such risk mitigation can prevent the expected loss of at most $[1-p]V_m$ of the contract value to the customer, as the expected loss of the service portion of the contract value to the customer, $[1-p]V_s$, is unrecoverable. This results in a value

wedge or deadweight loss in terms of economic surplus. To ensure that the surplus related to this part of the contract value is not lost, the intermediary has to be solvent at $t = T$.[16] We will show later that even if the customer could purchase a guarantee that compensates the customer for all of the service value of the contract, the customer's loss in utility from being exposed to the intermediary's credit risk is not lessened. Second, this suggests that the more efficient solution is for the *intermediary* to undertake risk mitigation to insulate the contract from its own credit risk, rather than expect the customer to do it. Merton (1997) identifies various ways in which the intermediary can do this; we take up this issue in Section 4.

It is important to note that this result does not depend on risk aversion, in the traditional sense, on the part of the customer. Risk aversion may be one particular way to capture this phenomenon. But if one resorts to this explanation, then it should be emphasized that this would be risk aversion *with respect to the uncertainty about the ability of the intermediary to deliver the embedded promise in the contract itself*, and not necessarily the randomness in the final payoffs that the contract might specify the customer would be exposed to. For example, a customer may indeed *expect* the final payoffs of the contract to be risky (as in a stock index mutual fund or a swap contract), but it is not risk aversion with respect to these payoffs that should play a special role in any explanation based on the risk aversion of customers. That is, the normal concept of risk aversion related to holding stocks and bonds does not accurately capture the behavior of customers that we are discussing here, where we are comparing the efficacy of alternative service-delivery contracts the customer has with the financial intermediary.

[16] Thus, another way of thinking about V in relation to the earlier discussion, is that V_m could be viewed as the monetary equivalent of the standard utility over wealth for risk-taking (i.e., the monetary flows that the contract stipulates is risky), while V_s can be viewed as a separate component for the services the intermediary provides, which the customer wants to be credit-insensitive.

3.2 The Inefficiency of Exposing Customers to Intermediary Credit Risk

We now provide a microfoundation for the idea discussed in the previous section that the customer should not be exposed to intermediary credit risk. This analysis should be viewed as a specific example of how economic surplus can be destroyed by exposing the customer to the intermediary's credit risk, but not the only way. In particular, we explain now why a simple resolution like having the customer buy an insurance contract that compensates the customer for all lost utility in the state in which the intermediary fails will not mitigate the inefficiency due to the customer's exposure to the intermediary's credit risk.

Consider a situation in which a financial institution raises financing from both customers and investors. The customers can be either risk averse or risk neutral. Since our focus is on the idiosyncratic credit risk of the institution, the assumption of investor risk neutrality is without loss of generality because they can diversify away the credit risk. So we will assume risk neutrality and a zero riskless rate. Financing from customers occurs because customers essentially "pre-pay" for future services, as described in the set-up in the previous section. For example, the customers of an insurance company purchase insurance and pay premia for possibly many periods before they experience an accident or some other contingency they have insured themselves against, a feature that is an essential element of the way insurance works and how insurance companies finance themselves. This timing of the service provision at a *future* date exposes customers to the institution's risk of failure.

For simplicity, we focus on two dates: $t = 0$ and $t = T$. The customer starts out at $t = 0$ with an endowment of $f > 0$ and all consumption occurs at $t = T$. The customer thus makes a single payment f to the intermediary at $t = 0$ and receives at $t = T$ a (possibly state-contingent) payment of F plus a bundle of services that yields the customer utility whose monetary value is u_s. These transfers

from the intermediary to the customer occur only if the intermediary is solvent.

Let the future state of the world at $t = T$ be represented by $\omega \in \Omega$, where Ω is the feasible set of states. In a subset $\Omega_1 \subset \Omega$ of states, the customer has a need for liquidity F that generates utility with a monetary equivalent of $u_l > F$ for the customer. In all other states, $\Omega \setminus \Omega_1$, the liquidity F provides a utility that has a monetary equivalent of F. In an insurance context, Ω_1 can be thought of as states in which there is an emergency need for funds for medical purposes, for example. In a banking context, this could be a set of states in which a proprietary positive-NPV opportunity is available. Let Ω_2 be a subset of states in which spot financing is unavailable to the customer to meet this liquidity need. This could be due to credit rationing (e.g. Stiglitz and Weiss (1981)) or because the customer has met with an accident that results in a disability that blocks access to credit. Finally, let Ω_3 be the set of states in which the intermediary fails, so F is not paid to the customer.

Now designate $\Pr(\Omega_1 \cap \Omega_2) = \xi$ and $\Pr(\Omega_1 \cap \Omega_2 \cap \Omega_3) = \xi[1 - p]$. As before, let k be the cost of intermediation to the intermediary. In the context of our previous discussion, we can write

$$V_s = u_s + \xi[u_l - F] \tag{4}$$

Note that in some situations, it may be that $u_s = 0$, and in other situations $u_l = F$. But at least one of the two components of V_s must be positive.

Theorem 2: *In the first-best case in which the institution faces no frictions in raising external financing at* $t = 0$ *and intermediation has social value, the contract between the institution and its customers completely protects customers from the credit risk of the institution related to its insolvency probability* $1 - p$.

The intuition is that there is a set of states in which the customer derives value from

intermediation services unavailable elsewhere, which in this specific construction is due to the customer deriving utility from the financing at *T* that in some states exceeds just the monetary value of the financing. This argument is preference-free, so it holds for risk-neutral as well as risk-averse customers. The corollary below now follows:

Corollary 1: *If there are two intermediaries with solvency probabilities p_1 and p_2 (with $p_1 > p_2$) and the market for intermediation services is perfectly competitive, so that prices adjust to solvency probabilities, then the customer will always strictly prefer the intermediary with p_1 to the intermediary with p_2.*

This corollary shows that simply adjusting the price of intermediation services to the solvency probability of the intermediary will not overcome the customer's aversion to the intermediary's credit risk. This is consistent with the Wakker, Thaler, and Tversky (1997) evidence discussed in the Introduction.

We want to stress that our notion of customer aversion to intermediary credit risk is broader than the specific construct that leads to Theorem 2 and Corollary 1. Nonetheless, these results are broad enough to cover many types of intermediary-customer contracts, such as those in banking and insurance.

The corollary above refers to a competitive market in which all the surplus goes to the intermediary's customers. We now examine what happens when intermediaries have monopoly power in customer markets.

Corollary 2: *If the intermediary is a monopolist in dealing with customers, the higher its solvency probability, the higher will be the surplus extracted by the intermediary from its relationship with the customer.*

This result shows that even a monopolistic intermediary will prefer to have a higher solvency probability because this will enable it to extract more surplus from the customer. That is, because customers enjoy a higher surplus at a higher intermediary solvency probability, the intermediary can extract more of this surplus.

3.3 Why is it Not Possible for Customers to Mitigate Intermediary Credit Risk?

A question one may ask at this stage is: why is it not possible to mitigate the welfare loss due to intermediary credit risk by simply creating a financial instrument that pays the customer enough money to offset the utility loss from the failure of the intermediary?

To address this question, note first that there are two ways in which one could attempt to do this. One is for the intermediary to purchase a third-party guarantee, as discussed by Merton (1997). We discuss this alternative later. A second way is for the customer to purchase this guarantee. We analyze that alternative here and show that it will *not* reduce the welfare loss due to the customer's exposure to the intermediary's credit risk.

A contract that completely protects the customer against intermediary default will pay the customer $F + u_s$ at $t = T$ if the intermediary defaults. Its cost at $t = 0$ will be $[1 - p][F + u_s]$. Since the customer's only endowment is f, this amount to purchase the guarantee will need to be borrowed.[17] Assuming that the customer must repay the lender at $t = T$ whenever the customer has enough resources to do so, it follows that this debt contract is riskless since the customer can repay the lender in all states. When the intermediary does not fail, the customer receives $F + u_s$ that can

[17] We assume a debt contract for financing the purchase, but our results do not depend on whether debt or equity is used.

be used to pay the lender.[18] When the intermediary fails, the guarantee pays $F + u_s$ that can be used to pay the lender. This leads to our next result.

Corollary 3: *The customer's borrowing money to purchase a guarantee against intermediary default has no impact on the loss in welfare due to intermediary credit risk.*

The intuition is that there is "no free lunch". A guarantee can protect against intermediary default, but the guarantee must be purchased via borrowing. Future repayment on the borrowing exactly offsets the benefit from the guarantee, leaving welfare unchanged.

There are also other ways in which customers may attempt to mitigate intermediary credit risk. Two other possible ways for customers to do this include: (i) diversifying across many intermediaries, or (ii) accessing an Arrow-Debreu market in primitive state securities to replicate the vector of services provided by the intermediary without being exposed to the credit risk of the intermediary. We explain now why both are either inefficient or infeasible.

First, consider (i). To diversify away the intermediary's idiosyncratic credit risk, the customer would have to replace its single-intermediary contract with a large number of smaller contracts with many intermediaries. However, one reason why we have financial intermediaries is that they achieve economies of scale and scope and reduce transaction costs; in our model, this would be reflected in k being, say, invariant to or concave in the size of the intermediary's contract with the customer. Thus, any attempt on the customer's part to diversify across intermediaries will be inherently inefficient due to duplicated costs of information acquisition and service provision.

Now consider (ii). Our argument is that replicating services is often infeasible because of market incompleteness in contracting. The incompleteness is that the customer cannot typically

[18] We are assuming that u_s is transferrable from the customer to the lender. This will often not be possible. Accounting for this non-transferrability complicates the analysis but does not change our results.

purchase a separate (Arrow-Debreu) claim that would deliver the service that the intermediary provides when it is solvent. That is, the intermediary is unique in providing its service once it has entered into a contract with the customer. In a complete market, the monetary and service components of the intermediary's contract would be traded separately as bundles of primitive Arrow-Debreu claims. This would enable the customer to purchase market-based insurance against the intermediary's credit risk. However, our analysis in Corollary 2 shows that even in this case, the welfare loss due to intermediary credit risk cannot be reduced. But the more realistic situation is that it is often physically impossible to purchase such insurance because the service the intermediary provides is typically inseparable from the monetary component of the contract, as explained in Section 3.1.[19]

Even if physical separability of these two components was possible, markets for intermediary services to customers would not be complete because the service is something that has to involve a contractual *relationship* between the intermediary and the customer—it cannot be something remote from the intermediary that can be traded in an anonymous market and purchased by the customer. This essential coupling of a specific intermediary with a specific customer often generates valuable customer-specific information that is available privately only to the intermediary, information that the intermediary can use to enhance the value of its service to the customer, as suggested by the relationship banking literature, for example. This rules out a complete market in which state-contingent claims can be created with values that depend only on states of the world and not on the "institutional affiliation" of each claim.[20]

[19] Moreover, the possibility of purchasing such primitive claims (with no contract risk) to replicate the desired payoff would mean that there would be no economic role for the financial intermediary in the first place.

[20] This is somewhat similar to the idea in Froot and Stein (1998) that financial institutions hedge the risk of illiquid assets in the capital market.

4. Financing Frictions and the Second-Best (Constrained-Efficient) Contract

In our discussion of the first-best case, we assumed that the intermediary faced no financing frictions. In the absence of such frictions, all of the intermediary's credit risk is efficiently borne by its investors and none by its customers. But we know from Myers and Majluf (1984) that adverse selection can make external finance costly relative to internal finance. In this section, we discuss the implications of this financing friction for the extent to which intermediaries choose to protect their customers from their own credit risks. That is, we analyze how financing frictions can cause the second best to deviate from the efficient (first-best) contract.

4.1 Financing Frictions and the Second Best

To see the effect of financing frictions, consider the three approaches suggested by Merton (1995) that financial intermediaries could use to protect their customers against intermediary credit risk. The first of these is for the intermediary to match its asset and liability payouts. While many intermediaries do attempt to reduce the maturity gap between their assets and liabilities, they typically do not eliminate this gap. A key reason for this is that maturity transformation is an important economic function served by many intermediaries, so a maturity gap is linked to the *raison d'etre* of the intermediary.

The second approach is for the intermediary to protect its customers by reducing its credit risk through an increase in its equity capital on its balance sheet. However, there is now an extensive literature on the costs that higher equity can entail. This include adverse selection costs (Myers and Majluf (1984)), loss of debt tax shields, and loss of debt discipline (e.g. Hart (1995)), among

others. These costs may limit the intermediary's reliance on equity capital to reduce credit risk.

The third approach is for the intermediary to purchase a guarantee from a credible third party. This approach generates a cost, however, when there is moral hazard because the intermediary can choose a (privately costly) hidden action. In the absence of a guarantee, investors will reflect this moral hazard in the pricing of securities.[21] This pricing is a source of capital market discipline, and it can reduce value-depleting actions the intermediary may choose. Examples of such discipline include the actions creditors can take--monitoring, maturity shortening in response to increased credit risk, and restrictions imposed on the intermediary when covenants are violated. In what follows, we present a formal example of how the moral hazard associated with a third-party guarantee in the second best can cause the customers of the intermediary to be exposed to risk.

However, this investor discipline can be supplemented with discipline imposed by the intermediary's *customers*. We argued in the previous sections that customers will display extreme aversion to being exposed to the intermediary's credit risk. This aversion too can be a source of market discipline if customers are even partially exposed to the intermediary's credit risk. That is, if the intermediary does not devote enough resources to significantly reducing its credit risk and thus exposes customers to it, these customers will flee the intermediary, as discussed in the Introduction (see Merton (1997)).[22] This threat of loss of customers can provide additional discipline on the intermediary. By purchasing insurance from a guarantor or by hedging, the intermediary could choose to completely insulate the customer from its own credit risk, but then market discipline from customers would be lost. So the second-best contract may involve

[21] Examples of such hidden action are the effort choices of the intermediary's managers, the risk profiles of the projects the intermediary invests in, the resources it devotes to risk management, etc.

[22] This includes instances in which this credit risk exposure results from inefficient choices by the bank. Calomiris and Kahn (1991) model such a situation. In their model, (uninsured) depositors flee the bank if they observe/suspect that the bank manager is making bad investment decisions, and this threat disciplines the manager.

customers being (at least potentially) exposed to some credit risk.

4.2 A Formal Analysis of CCF Costs in the Second Best

We lay out here a simple theoretical model to show how CCF costs can arise in the second best. We will also use the setup of this model in Section 6 to examine the growing integration of banks and financial markets.

Consider a commercial bank that has access to uninsured deposits at $t = 0$. It raises f in deposit funding at $t = 0$ and promises its depositors (customers) a repayment of F at $t = 1$. The banks invests the f it raises from depositors to lend to a borrower at $t = 0$. The loan pays off a random amount $\widetilde{X}$ to the bank at $t = 1$, which is $X \in \mathbb{R}_+$ with probability $p(e) \in [0,1]$ and zero with probability $1 - p(e)$. Here $e \in [0,1]$ is the bank's monitoring effort chosen at $t = 0$, with $p(0) = 0$, $p' > 0$, $p'' \leq 0$. That is, consistent with the relationship banking literature (e.g. Boot (2000), and Holmstrom and Tirole (1997), we assume that the bank's loan monitoring increases the probability of loan repayment. The cost (stated in terms of value at $t = 1$) of the effort to the bank is $\varphi(e)$, $\varphi(0) = 0$, $\varphi(e) > 0 \; \forall \; e > 0$, $\varphi' > 0$, $\varphi'' > 0$, $\varphi''' > 0$, and the Inada conditions $\varphi'(0) = 0$, $\varphi'(1) = \infty$. To this, we add the technical condition $\varphi'' > V_s \; \forall \; e > 0$, where V_s is the service value received by depositors. The monitoring effort e is unobservable to all except the bank and hence cannot be directly contracted upon.

All agents are risk neutral and the riskless interest rate is $r - 1$ (and therefore the discount factor is r^{-1}). The bank's loan is efficient in the sense that:

$$r^{-1}p(e)X - f - \varphi(e) > 0 \; \forall \; e > 0 \tag{5}$$

Because the loan is risky, the uninsured depositors are exposed to the bank's credit risk. Suppose the bank can purchase from a third-party guarantor a competitively priced (to yield the guarantor

an expected return of $r - 1$) guarantee that pays the bank a fraction $\gamma \in [0,1]$ of X in the state in which the borrower defaults. We assume that the fraction of the service value, V_s, that depositors are able to enjoy is 1 if the bank realizes its full payoff X, and $\lambda \in [0,1]$ if the bank realizes less than X. That is:

Fraction of X that Guarantor Pays Bank when Borrower Defaults (γ)	Fraction of V_s enjoyed by depositors (λ)
$< \gamma_0$	0
γ_0	$\lambda_0 \in (0,1)$
1	1

Here, $\gamma_0 \in [\gamma_1, \gamma_2]$, where $\gamma_1 X = \bar{\gamma} F$ for some exogenous $\bar{\gamma} \in (0,1)$ and $\gamma_2 X = F$. As a notional matter, for any fraction γ of X that the guarantor pays the bank, define $\tilde{\gamma}$ as the corresponding fraction of F that the payment corresponds to. Clearly, $\tilde{\gamma} \geq \gamma$.

Let us now interpret the specification above. The basic idea is that in order to provide V_s to the depositors, the bank has to commit internal resources. If the guarantor pays the bank in the state of borrower default what it would have received had the borrower repaid, then the bank is essentially solvent despite borrower default and hence the depositors enjoy the full value of services, V_s. If the bank recovers less than X from the guarantor, then the bank is bankrupt and its ability to deliver depository services is impaired (similar to Bernanke's (1983) argument that bankruptcy impaired banks' lending capabilities). For all values of γ such that depositors are repaid a fraction ranging from $\bar{\gamma}$ to 1 of what they are owed, depositors enjoy a fraction λ_0 of V_s. This means that even if depositors are fully repaid (F), they may not recover the full value of their services if the bank is bankrupt, consistent with the idea mentioned in Section 3 that customers cannot recover the full value of V_s by hedging their financial claim. And if depositors are repaid a

fraction less than $\bar{\gamma}$ (i.e., the bank receives less than a fraction γ_0 of X), then the bank's service capability is so impaired that depositors lose all of the service value, V_s.

For the bank, γ is a choice variable. That is, the bank solves:

$$\max_{\gamma \in [0,1]} [r^{-1}\{p(e)[X-F] + [1-p(e)]I_0[X-F] - k - \varphi(e)\} - g] \tag{6}$$

subject to

$$e \in \arg\max_{e \in [0,1]} \{r^{-1}[p(e)[X-F] + [1-p(e)]I_0[X-F] - \varphi(e)\}] \tag{7}$$

$$f = r^{-1}\left[\begin{array}{c} p(e)[F+V_s] + [1-p(e)]I_0[F+V_s] + [1-p(e)]I_1[\tilde{\gamma}F + \lambda_0 V_s] \\ + [1-p(e)]I_2\tilde{\gamma}F \end{array}\right] \tag{8}$$

$$g = r^{-1}[1-p(e)]\{I_0X + I_1\gamma_0X + I_2\gamma X\} \tag{9}$$

where $I_0 = 1$ if $\gamma = 1$ and 0 otherwise; $I_1 = 1$ if $\gamma = \gamma_0 \in [\gamma_1,\gamma_2]$ (i.e., if $\tilde{\gamma} \in [\bar{\gamma},1]$) and 0 otherwise; and $I_2 = 1$ if $\gamma < \gamma_0$ and 0 otherwise. Moreover, k is the bank's cost of providing depository services, and g is the cost of the guarantee to the bank at $t = 0$. Here, (6) is the bank's maximization of the value of its equity, (7) is the Nash constraint on the bank's choice of effort, (8) is the deposit pricing constraint, and (9) is the competitive-pricing determination of g.

We now have:

Lemma 1: *The first best solution involves an effort choice* $e^0 \in (0,1)$ *by the bank and* $\gamma = 1$.

The intuition for $\gamma = 1$ is straightforward. Any guarantee less than that sacrifices some of the depository services surplus, V_s, enjoyed by depositors, which is inefficient. Thus, in the first best, not only is the depositor' financial claim completely covered, but so is the value of the services they receive from the bank. We will now show that such complete protection is impossible in the second best when e cannot be contracted upon.

Theorem 3: *In the second best,* $\gamma = \gamma_0$.

The intuition is as follows. Implementing the first best is impossible when e is unobservable because the bank's payoff becomes insensitive to its choice of e, so it chooses $e = 0$. Choosing $\gamma < \gamma_0$ is inefficient as well because all of the service value depositors enjoy, V_s, is lost if the borrower defaults on the bank loan. Choosing $\gamma = \gamma_0$ ensures that the bank's shareholders receive nothing when the borrower defaults, so the strongest second-best effort incentive is maintained, with the highest fraction of V_s made available to the depositors. Note that with $\gamma = \gamma_0$, the depositors receive a financial payoff between $\bar{\gamma}F$ and F and enjoy $\lambda_0 V_s$ in terms of the service value of deposits. That is, they are exposed to the bank's credit risk. If we added a small adverse selection cost associated with the purchase of the guarantee, then the bank will purchase a guarantee that pays depositors $\bar{\gamma}F$ when the bank is bankrupt. In any case, in the second best, depositors face exposure to the bank's credit risk, either for only the service value of their deposits or for the service value as well as the financial claim.

4.3 Examples of Practical Solutions to the Second-Best Exposure of the Customer to Intermediary Credit Risk

Banking regulators seem to be aware of the potential efficiency losses from the exposure of the bank's customers to intermediary credit risk in the second-best case. Several pieces of international regulation are intended to lessen these efficiency losses (e.g. Vickers in the U.K., Liikanen in the EU). The Vickers Report (Vickers (2011)) requires banks to ring-fence their retail banking operations from the riskier wholesale and investment banking operations by organizing

its retail banking operations in a heavily-regulated retail banking subsidiary.[23] Similarly, the Liikanen Report (Liikanen (2012)) proposes walling off in a separate subsidiary the bank's trading activities from the bank's deposit taking and retail-payment operations. In the U.S., the "living wills" in the Dodd-Frank Act are intended to reduce perceived counterparty risk and facilitate the continuation of certain activities; see also our discussion of Dodd-Frank in Section 5.

Of course, banks themselves sometimes engaged in ring-fencing even without regulatory prompting. An example is the case of Salomon Brothers and RJR Nabisco. When Salomon expressed interest in undertaking a leveraged buyout of RJR Nabisco in 1988, many of Salomon's credit-sensitive customers fled because of concerns over how this would affect its overall creditworthiness. Salomon's response to mitigate this concern in subsequent years was to create a "ring-fenced" AAA-rated subsidiary called Salomon-Swapco as a counterparty for its OTC customers' derivatives trades.

5. Examples of Customer Contracts, Institutional Design, and Regulatory Practices

In this section, we discuss how our analysis can shed light on some observed financial contracts, institutions, and regulatory practices. We end the section with a discussion of which products and services are the most sensitive to the losses from intermediary credit risk and hence require government intervention, which then leads to a discussion of the role of the government in the financial sector.

[23] See Greenbaum, Thakor, and Boot (2015) for details.

5.1 Customer Contracts

5.1.1 Bank Deposits

A demand deposit in a bank represents a contract between the bank and a customer (depositor). In practice, the funds provided by depositors are invested in risky securities (e.g. loans), and uninsured depositors are exposed to the credit risk of the bank, consistent with the second-best contract. However, the depositor would prefer not to have to worry about the credit risk of the bank if the bank could find a cost-effective way to achieve this.

There are many ways for a bank to protect its depositors against its own credit risk, although they all entail potential costs. For example, narrow banking, whereby a bank can invest only in safe assets such as U.S. Treasuries, would eliminate bank credit risk but it would require that the bank abandon its key economic services in loan origination, screening, and monitoring; this would represent a potential economic loss. Similarly, requiring that the bank put up a substantial amount of equity may also entail significant costs, depending on the magnitude of the equity infusion, as discussed earlier. Deposit insurance is a solution that avoids those costs, and our analysis offers a rationale for deposit insurance that does *not* rely on preventing runs.[24] Even if contagious bank runs are not a problem, deposit insurance improves efficiency in our theory because it enables one to move closer to the first best in which the bank's customers are completely insulated from its credit risk. Deposit insurance is one reason why customers (retail depositors) are willing to deal with institutions that lack a AAA credit rating.[25] The fact that depositor insurance is incomplete

[24] Preventing runs is the most widespread justification for deposit insurance in the literature, dating back to Bryant (1980) and Diamond and Dybvig (1983).

[25] These ratings refer to the credit risk to which the bank's uninsured creditors (investors) are exposed, not its depositors. Deposit insurance makes the deposit contract riskless for the financial claim of the insured (core) depositors—even if the bank lacks a AAA rating on its uninsured debt. These core depositors represent the bulk of the customers our framework focuses on, with perhaps a small fraction of these customers being partially insured and representing a source of market discipline.

may be understood in the context of the need for customer-imposed market discipline to deal with moral hazard.[26]

Our analysis also has implications for proposals like "bail-in deposits" that have been put forth as a way to infuse more equity capital into distressed banks (see Zenios (2016)). The idea is these deposits would convert to equity in a distressed bank. Our analysis says that this may be inefficient if the depositors are customers, and would only make economic sense for those depositors (like some uninsured jumbo CD holders, for example) who are investors or expect to become investors.

5.1.2 Mutual Funds

For mutual funds, customers are investors in the fund—each customer is purchasing a service (the portfolio management service and the promise of some risky return), while also providing financing. In this case, the customer understands that the contract purchased from the mutual fund may have a risky payoff, for example, linked to the S&P 500. It is only the credit risk of the intermediary—say, due to unobserved risky investments with fund money or "tunneling"—that the customer wishes to be insulated from.

A mutual fund is a good example of a contract that imposes risk on the investor (customer), but only that risk which is confined to the contract itself.[27] Indeed, this is one reason why investors put their money in funds managed by reputable intermediaries like Vanguard, Fidelity, and the like. If one invests in the S&P 500 through one of these funds, the risk (R) is $R_{S\&P\ 500}$. Other than differences in expenses, the risk in the fund is the same regardless of whether it is offered by Vanguard or T Rowe Price or American Century. If these funds were rated, they would all be AAA, even though their future value is (systematically) risky. However, if one chooses to invest in the S&P 500 through an individual/company of lesser reputation, say agent *XYZ*, then the

[26] Merton (1977) shows how deposit insurance is isomorphic to a put option, and how it can create moral hazard.
[27] E.g. the custodian holds securities and provides insurance on theft.

investor's risk is $R_{S\&P\ 500} + R_{XYZ}$, where R_{XYZ} is the credit risk of *XYZ*. Thus, in the case of many mutual funds, the second-best contract closely approximates the first best as customers are exposed to little, if any, credit risk of the intermediary offering the fund.

5.1.3 Insurance Contracts

An individual who purchases a whole life insurance policy is a customer who is buying a bundle of two products—an insurance payoff in the event of death and an investment. The policyholder is willing to accept randomness in the return on the investment portion of the product, but not the risk that the insurance company may fail due to other exposures and hence be unable to pay in the event of the death of the insured. If the insured were to be exposed to such risk, it would represent a risk-sharing distortion because it would not be efficient for the insured to buy a large number of smaller life insurance policies to diversify across insurance companies, as we showed earlier.

A similar argument holds for property and casualty insurance. Our analysis in the previous section shows why the first-best insurance contract completely protects the customer from the credit risk of the insurance company. To the extent that the second best may leave the customer with some exposure, an insurance fund that backs up insurance companies and protects policyholders can enhance efficiency. In all fifty U.S. states, state insurance funds provide this service.[28] This moves the second-best contract closer to the first best.

5.1.4 Repurchase Agreements (Repos)

Repos have been the mainstay of short-term financing in the shadow banking sector for over a decade, and this sector is now globally bigger than commercial (deposit-based) banking. A repo

[28] This also applies to property and casualty insurance. Property and casualty guaranty funds are part of a non-profit, state-based system that was created by statute, which pays outstanding claims of insolvent insurance companies. As of 2015, there were about 550 insolvencies since the inception of the guaranty funds.

contract is an excellent illustration of our theory. The financial intermediary is an institution that has collateral in the form of bankruptcy-remote securities like U.S. Treasuries or high-grade mortgage-backed securities, but has need for liquidity over a short time period. The customer is another institution that has excess liquidity on which it wishes to earn additional yield income. The customer provides financing to the intermediary in exchange for taking ownership of the collateral for the duration of the loan. Since the loan amount is less than or equal to the value of the securities used as collateral, the customer is *not* exposed to the credit risk of the intermediary. As a result, the second-best arrangement approximates the efficiency of the first best.[29] In fact, the efficiency of the first best is achieved by the central bank repo facility (known as RRP) started by the Federal Reserve in 2013. This is an overnight repo program in which institutions are able to park cash with the Fed and earn a modest interest rate in exchange for U.S. Treasuries as collateral. Our theory helps to explain the growing popularity of this facility.[30]

5.2 Institutional Design: Futures Exchanges

A futures contract essentially guarantees the ability to sell or buy some commodity or security in the future at a price that is predetermined. If this contract were negotiated as a forward contract with a financial intermediary, the holder of the contract would be the intermediary's customer.[31] Clearly, if the intermediary becomes insolvent prior to the delivery or execution date on the

[29] Indeed, concerns about the extent to which the repo contract is insulated from the credit risk of the borrower can lead counterparties to refuse to enter into the contract or substantially increase the repo "haircut". For example, as it approached insolvency, Bear Stearns was unable to find repo counterparties with even Treasuries as collateral.
[30] See Burne (2016).
[31] With swaps and forwards, the two parties can switch back and forth between being a creditor or debtor to one another based on movements in the underlying assets. So approximately half the time, the customer will be owed money by the intermediary and thus be credit-sensitive. As noted previously, if a customer owes money to the intermediary, then he does not fit the definition of credit-sensitive customers that we focus on here. In this sense, options traded on an exchange may be a better example. However, even for swaps/forwards, as long as there is a chance that the customer will become the creditor in the contract (if the market value of the underlying goes the other way), the effect that we emphasize will still be present.

contract, the customer would be unable to avail of the insurance against the price risk that the customer sought under the contract.

A futures contract is traded on an exchange with liquidity and collateral provided daily, rather than being merely a bilateral arrangement between the bank and the customer that may not be collateralized. The exchange stands behind the execution of the contract. Consequently, the customer is protected against counterparty risk. Thus, the use of futures contracts over forward contracts may be rationalized as a means of insulating customers against the credit risk of an intermediary.

5.3 Regulatory Practices: The Dodd-Frank Act

One aspect of the Dodd-Frank Act that can be understood within the context of our framework is that under Title VII of the Act, all non-exempt swaps to which a clearing exception does not apply (i.e. “standardized” swaps) must be cleared and exchange traded. Mandatory clearing and exchange trading of swaps is already underway. Our analysis provides an economic rationale for this. By making swaps exchange-traded, counterparty credit risk is greatly reduced, moving the arrangement closer to first best. Thus, the customers who hold these swap contracts need not worry about the credit risk of the intermediary they are working with, provided that the exchange is bankruptcy remote. Thus, this requirement of Dodd-Frank serves the economic purpose of minimizing customer-specific contract fulfillment risk in swaps.

An interesting question is why market participants did not do this on their own prior to Dodd-Frank. There are numerous possible explanations for this. One reason may be coordination failures among market participants in the process of setting up an exchange. The fact that coordination failures can lead to adoption externalities that cause individual participants to avoid welfare-

enhancing initiatives has been established in various contexts.[32] For example, Dybvig and Spatt (1983) develop a model which explains that the failure to adopt the metric system in the U.S. may be attributable to such a coordination failure. Fishman and Hagerty (2003) develop a model in which mandatory information disclosure is rationalized on the grounds that voluntary disclosure may not be forthcoming when the fraction of consumers who can understand a disclosure is too low. These are examples of economic settings in which a regulatory mandate helps to overcome the negative adoption externalities generated by coordination failures.

5.4 A Perspective on the Role of the Government

Our theory provides a perspective on when the government should intervene in the financial sector. The simple idea is for the government to intervene when the second-best generates the biggest deadweight losses (CCF costs) from customer exposure to intermediary credit risk. In the context of the specific examples discussed in this section, deposits and insurance contracts would be at the top of the list in this regard. Mutual funds, repos, and futures exchanges would be at the bottom of the list; these are examples where market-mediated solutions are effective in generating low CCF costs, making government intervention unnecessary. Swaps exchanges would be somewhere in the middle of the spectrum.

This perspective on the role of the government also illuminates the regulatory practice of protecting the largest banks in an economy by considering them "too big to fail" (TBTF).[33] According to our theory, there are two reasons for this. One is that bigger banks are more complex

[32] Coordination failures in our context can take many forms. For example, low-default-risk counterparties who are privately informed about each other's credit risk may prefer to engage in off-exchange bilateral contracts and avoid exchange-trading costs. This may result in a downward quality spiral in firms that voluntarily choose exchange trading, a problem that is potentially exacerbated by higher costs of setting up an exchange, which may be the case with highly-customized swap contracts. This can adversely affect liquidity in exchange trading.
[33] See Kaufman (2013) for a discussion of too-big-to-fail.

than smaller banks, with potentially greater *intertwining* of customers and investors. The bigger the bank, the more difficult it becomes to keep investors separate from customers. Moreover, even apart from this, large banks may be more interconnected (see Allen, Babus, and Carletti (2012)), so it may be more difficult to determine whether customers have been effectively insulated from the bank's risk. Consequently, the second-best contract between the bank and its customers exposes customers to more credit risk than in less complex organizations, with higher CCF costs ex ante and greater damage to customers ex post if the bank fails.

The other reason is that if large banks fail, they may have to sell so many assets that a fire sale may ensue. This would imperil other (smaller) banks that would experience asset value declines and possible failure due to mark-to-market accounting, which in turn would hurt the customers of those banks. For example, the Federal Reserve stepped in to assist Bear Stearns when it was on the verge of collapse in 2008 because of its interconnectedness and a concern about the possible effect of its failure on the customers of the investment bank; it was involved in trillions of dollars of repo agreements and swap contracts that made it interconnected with many institutions and also raised the specter of a possible fire sale if the bank failed and replacement counterparties could not be found. In the process of protecting its customers, of course, the investors of Bear Stearns were also protected. For these two reasons, TBTF may make economic sense because it can more effectively protect customers and reduce CCF costs.

6. The Growing Integration of Banks and Financial Markets

The 2007-2009 financial crisis showed how integrated the depository institutions are with financial markets. This integration blurs the boundary between banks and markets and complicates bank regulation and the government's approach to crisis resolution. For example, one of the

reasons why the Federal Reserve assisted in the acquisition-based rescue of Bear Stearns was that the investment bank was a counterparty in trillions of dollars of swap transactions and there was a concern about the potential consequences of not enough institutions stepping in to replace Bear Stearns if it failed. The purpose of this section is to examine how the extent and nature of this integration are affected by the bank-customer relationship in our framework.

This integration has occurred in a variety of ways. These include loan originations by banks that end up creating asset-backed securities through securitization that are then sold to investors in the capital markets, including other banks; market-traded credit-default swap (CDS) contracts that insure against default by bank borrowers; the use of exchange-traded derivatives of various sorts that are used in conjunction with more customized non-traded derivatives created by banks to reallocate risks; and loan commitments sold by banks to borrowers as lines of credit to back up commercial paper issues sold in the market. A number of authors have discussed this blurring of the boundary between banks and markets, why it is occurring, and its ramifications.[34]

We address a different question here related to this issue: what determines the *extent* to which banks will choose to integrate with markets? That is, how does our functional perspective shed light on the dynamics of the evolving boundary between banks and markets? To examine this question, we deploy a simple extension of the model developed in Section 4. Below, we describe only the portions of the model that differ from those in Section 4.

The analysis below leads to two main conclusions. First, although the deposit insurance safety net encourages the bank to take market integration risk, the sensitivity of the bank's customers to this risk provides a counterbalance. Second, the bank will only engage in capital market trading activities that have higher expected profit than lending *in equilibrium*, taking into account the

[34] See, for example, the various papers in Berger, Mollyneaux, and Wilson (2011).

bank's loan monitoring.

Consider a commercial bank that has access to (completely) insured deposits at date $t = 0$; so a third-party guarantee is unnecessary. It raises f in deposit funding at $t = 0$ and promises its depositors (customers) a repayment of F at $t = 1$. The monetary component of what the depositors receive from the bank is thus F, regardless of the bank's fortunes, due to deposit insurance. However, suppose the depositors do not receive the service component of the deposit contract if the bank fails and the deposit insurer has to pay depositors, i.e., this corresponds to $\lambda_0 = 0$ in Section 4. The deposit insurance premium, π, is set to be actuarially fair. Assumptions on preferences and the riskless rate are the same as in Section 4.

The bank uses the f it raises at $t = 0$ from depositors to lend to a borrower and to invest in market-based activities that increase the bank's integration with the market.[35] We assume that a fraction $\alpha \in [0,1]$ of f is invested in the loan and the remainder, $1 - \alpha$, is invested in market-based activities.[36] We can view $1 - \alpha$ as a measure of the "integration" of the bank with the market. For every dollar invested at $t = 0$, the loan yields a random repayment of $\tilde{x}$ at $t = 1$, where $\tilde{x} = x \in \mathbb{R}_+$ with probability $q + p(e)$, and 0 with probability $1 - q - p(e)$, where $q \in (0,1)$, $p(e) = e[1 - q]$, and $e \in [0,1]$ is the bank's privately-observed monitoring effort.[37] The cost of the effort to the bank is $\varphi(e)$, which has the properties described in Section 4.

For every dollar invested in the bank's market-based activities at $t = 0$, the payoff at $t = 1$ is a random variable $\tilde{z}$ which takes a value of $z \in \mathbb{R}_+$ with probability q and 0 with probability $1 - q$. We assume that $z > x$, but $x > qz$ (absent this, the bank would never invest in the loan). We assume:

$$qx > r \tag{10}$$

[35] We exclude bank equity for simplicity because it plays no role in the analysis.
[36] These could be proprietary ("prop") trading, writing CDS contracts, purchasing traded securities for investment, etc.
[37] Since $\tilde{x}$ is a per-dollar-invested payoff, we can think of $\alpha f \tilde{x} = \tilde{X}$ in the context of the notation in Section 4.

so both the loan and the market-based investments are efficient, but absent bank monitoring effort, market-based activities are more profitable (since $z > x$). For simplicity, $\tilde{x}$ and $\tilde{z}$ are orthogonal random variables, and we assume that the bank cannot fully repay depositors unless it experiences success on its loan. The idea is that we are considering a bank with most of its assets in loans (high α) and examining its decision to allocate some capital to market-based activities.

This specification means that the bank can reduce its loan default risk by monitoring, but cannot influence the risk of its market-based activities. The bank thus faces a tradeoff. On the one hand, the more it invests in market-based activities, the greater is the elevation of its expected profit without having to incur monitoring costs. On the other hand, investing in lending allows the bank to reduce its risk of failure through (costly) monitoring. Even though all agents are risk neutral, the bank cares about its risk of failure because it affects the value of the depository services it offers its customers, and this affects the extent to which the bank integrates with the market.

The sequence of events is as follows. The bank chooses $\alpha \in [0,1]$ at $t = 0$, which then determines its monitoring effort e. After that the depositors observe α and determine F, given the f that the bank seeks to raise at $t = 0$. Repayment from the borrower to the bank and the bank's payoff on its market-based activities both occur at $t = 1$. The bank chooses α to maximize the value of the bank to its owners at $t = 0$. That is, the bank solves:[38]

$$\max_{\alpha \in [0,1]} r^{-1}\{\alpha fx[q + p(e)] + [1 - \alpha]fqz - k - F - \varphi(e)\} \tag{11}$$

subject to

$$\operatorname*{argmax}_{e \in [0,1]} r^{-1}\{[\alpha fx - F][p(e) + q] + [1 - \alpha]fzq[p(e) + q] - \varphi(e)\} \tag{12}$$

[38] The bank's objective function in (11) is actually $\max_{\alpha \in [0,1]} r^{-1}\{[\alpha fx + [1 - \alpha]fz - F][p + q]q + [\alpha fx - F][p + q][1 - q] - k - \varphi(e)\} - \pi$, where $\pi = r^{-1}\{[1 - p - q]q[F - [1 - \alpha]fz] + [1 - p - q][1 - q]F\}$. Substituting from π back into the objective function yields (11).

$$f = r^{-1}[F + \{q + p(e)\}V_s] \tag{13}$$

Recall that k is the intermediation cost of service provision by the bank and V_s the value of these services to depositors, consistent with the notation that was introduced earlier.

In the above, (11) is the net wealth of the bank's shareholders and the maximization with respect to α assumes e is optimally chosen (12) is the (Nash) constraint that determines the bank's monitoring effort, and (13) is the competitive deposit pricing constraint. Note that the $p(e)$ in (8) is based on the depositors' belief about the equilibrium e^* that the bank will choose (a belief that must be correct in equilibrium). The depositors will base their beliefs on the α the bank chooses, and the bank will account for this inference in setting α.

Lemma 2: *The bank's optimal monitoring effort, e^*, is an interior optimum satisfying $de^*/d\alpha > 0$ and $d^2e^*/d\alpha^2 < 0$ for any given α.*

The interior optimum for monitoring effort results from the convexity of φ and the Inada conditions. What this lemma tells us is that the bank's optimal monitoring effort is increasing and concave in α, the bank's investment in lending. This is intuitive—a greater allocation of capital to lending increases the bank's marginal return to loan monitoring effort. Next we have:

Theorem 4: *There is an interior optimum $\alpha^* \in [0,1]$ that is a solution to the program in (6)-(8). The higher is the per-dollar payoff, z, from a successful capital-market investment, the lower is α^*. Moreover, the equilibrium choice of α^* always ensures that the per-dollar expected payoff on market financing exceeds the per-dollar expected payoff on the loan.*

The intuition is as follows. As capital-market activities become more profitable for the bank, integration becomes more attractive, facilitated by the bank's access to insured deposits. However,

since depositors attach value to depository services, their "sensitivity" to the bank's credit risk makes it less attractive for the bank to integrate with the market. So in equilibrium, market-based activities have to be more profitable for the bank than lending.

One key assumption in our analysis is that the bank's failure will disrupt (liquidity) service provision to depositors. This is what makes the second best deviate from the first best (in which case V_s would also be insulated from the bank's credit risk). If this were not the case, there would be no "market discipline" on the bank to rein in its integration. Regulators would then need to rely increasingly on regulations like the "Volcker Rule", with its prop trading proscription, which can be viewed as an attempt to limit bank-market integration.

Regulatory costs can also play a role in how much banks choose to integrate. If regulatory costs or the cost of monitoring bank loans go up significantly, banks may find it optimal to reduce V_s because CCF costs are higher. This will have both a direct and an indirect effect on the extent to which banks will choose to integrate with markets. The direct effect is that, holding *f* fixed, the bank will choose a lower α^* when V_s declines. The indirect effect is that *f* may fall as customers move from bank deposits to mutual funds as higher relative rates of return there compensate partially for the decline in utility due to the loss of bank-deposit-related services. The bank will replace lost deposits with investor funding through subordinated debt, for example. This makes V_s less important to the bank and decreases α^*. Thus, both the direct and indirect effect are predicted to increase the integration between banks and markets when banks' regulatory (or loan monitoring) costs go up. This will further lower e^*, reduce the competitiveness of banks in the loan market, and increase the banks' customers' exposure to counterparty (bank credit) risk, a sort of "multiplier" effect. The greater integration also means that market risks will impact the fortunes of banks more, creating an unintended *de facto* expansion of the government safety net as non-

depository market institutions that affect the risks of insured banks are bailed out in an effort to limit the likelihood of (even bigger) insured banks failing.

7. Conclusion

In this paper, we have developed the notion of "customers" and "investors" as a framing of the roles played by different groups of agents in funding financial intermediaries. Customers provide a significant amount of the funding, but want to bear no intermediary credit risk. In contrast, investors provide both funding and risk-bearing. The customers' dislike for the credit risk of the intermediary makes them different from investors, and this distinction leads to a rich set of implications.

The most important implication that we focus on in our framework is the economic rationale for designing efficient (first-best) contracts that insulate customers from the credit risk of the intermediary and impose all of this idiosyncratic risk on the investors. We show that because customers cannot replicate the services they receive from intermediaries due to a form of market incompleteness, exposing customers to the idiosyncratic credit risk of the intermediary results in an inefficient loss of economic surplus. Intermediaries thus have an incentive to design contracts that protect the customer from the intermediary's credit risk. However, in a second-best setting in which providing such risk insulation is costly, a tradeoff must be made between the intermediary-specific cost of insulating the customer from the credit risk of the intermediary and the cost of leaving the customer partially exposed. This perspective helps to explain the design of a variety of real-world contracts—including not only deposit contracts in banking, but also the other contracts such as mutual funds, insurance contracts, and repos. It also provides a fresh perspective on why deposit insurance may be efficiency-enhancing even in the absence of contagious runs, and why

we have securities exchanges. Moreover, it generates an economic rationale for the swaps clearinghouse requirement of the Dodd-Frank Act. An empirical implication of our theory is that whenever customers' concerns about the credit risks of the financial institutions they deal with are elevated, institutions that offer their customers better protection will gain a competitive advantage.

Our perspective in this paper also illuminates the blurring distinctions between banks and markets. Whenever the CCF costs involved in the provision of intermediated services become sufficiently high, or the value of the bank's depository services to its customer declines, banks increase their integration with the capital market, further blurring the boundary between the two.

We view this theoretical framework as a useful starting point for identifying and understanding how the key roles of customers and investors impact financial intermediaries. Future work could take the framework further, and examine the implications for regulatory policy involving systemic risks. In addition, our theory also has implications for how certain types of contracts should optimally be structured, such as debt contracts between intermediaries and customers. An interesting extension would be to look at how our approach bears on the work related to opacity and transparency in contracts.

References

1) Allen, Franklin, Ana Babus, and Elena Carletti. "Asset commonality, debt maturity and systemic risk." *Journal of Financial Economics* 104, no. 3 (2012): 519-534.
2) Allen, Franklin, and Douglas Gale. *Understanding financial crises*. Oxford University Press, 2009.
3) Berger, Allen N., Philip Molyneux, and John OS Wilson, eds. *The Oxford handbook of banking*. OUP Oxford, 2014.
4) Bernanke, Ben S. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *The American Economic Review* 73, no. 3 (1983): 257-276.
5) Boot, Arnoud WA. "Relationship Banking: What Do We Know?." *Journal of Financial Intermediation* 9, no. 1 (2000): 7-25.
6) Boot, Arnoud WA, and Anjan V. Thakor. "Security design." *The Journal of Finance* 48, no. 4 (1993): 1349-1378.
7) Bryant, John. "A model of reserves, bank runs, and deposit insurance." *Journal of Banking & Finance* 4, no. 4 (1980): 335-344.
8) Burne, Katy, "Fed Repo Program Swells", *The Wall Street Journal*, October 3, 2016. p. C7.
9) Calomiris, Charles W., and Charles M. Kahn. "The role of demandable debt in structuring optimal banking arrangements." *The American Economic Review* (1991): 497-513.
10) Campbell, Larry, and John P. Wilson, "Financial Functional Analysis: A Conceptual Framework for Understanding the Changing Financial System". Working paper, HSBC London, 2014.
11) Dang, Tri Vi, Gary Gorton, Bengt Holmström, and Guillermo Ordonez. "Banks as secret keepers." *The American Economic Review* 107, no. 4 (2017): 1005-1029.
12) DeAngelo, Harry, and René M. Stulz. "Liquid-claim production, risk management, and bank capital structure: Why high leverage is optimal for banks." *Journal of Financial Economics* 116, no. 2 (2015): 219-236.
13) DeMarzo, Peter M. "The pooling and tranching of securities: A model of informed intermediation." *Review of Financial Studies* 18, no. 1 (2005): 1-35.
14) DeMarzo, Peter, and Darrell Duffie. "A liquidity-based model of security design." *Econometrica* 67, no. 1 (1999): 65-99.
15) Diamond, Douglas W. "Financial intermediation and delegated monitoring." *The Review of*

Economic Studies 51, no. 3 (1984): 393-414.

16) Diamond, Douglas W., and Philip H. Dybvig. "Bank Runs, Deposit Insurance, and Liquidity." *The Journal of Political Economy* 91, no. 3 (1983): 401-419.

17) Dybvig, Philip H., and Chester S. Spatt. "Adoption externalities as public goods." *Journal of Public Economics* 20, no. 2 (1983): 231-247.

18) Fishman, Michael J., and Kathleen M. Hagerty. "Mandatory versus voluntary disclosure in markets with informed and uninformed customers." *Journal of Law, Economics, and organization* 19, no. 1 (2003): 45-63.

19) Froot, Kenneth A., and Jeremy C. Stein. "Risk management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach." *Journal of Financial Economics* 47, no. 1 (1998): 55-82.

20) Fulghieri, Paolo, and Dmitry Lukin. "Information production, dilution costs, and optimal security design." *Journal of Financial Economics* 61, no. 1 (2001): 3-42.

21) Gorton, Gary, and George Pennacchi. "Financial Intermediaries and Liquidity Creation." *The Journal of Finance* 45, no. 1 (1990): 49-71.

22) Greenbaum, Stuart I., Anjan V. Thakor, and Arnoud Boot, eds. *Contemporary financial intermediation*. Academic Press, 2015.

23) Hanson, Samuel G., Andrei Shleifer, Jeremy C. Stein, and Robert W. Vishny. "Banks as patient fixed-income investors." *Journal of Financial Economics* 117, no. 3 (2015): 449-469.

24) Hart, Oliver. "Corporate governance: some theory and implications." *The Economic Journal* 105, no. 430 (1995): 678-689.

25) Hart, Oliver, and Luigi Zingales. “Banks Are Where The Liquidity Is.” No. w20207. National Bureau of Economic Research, 2014.

26) Hirshleifer, Jack. "The Private and Social Value of Information and the Reward to Inventive Activity." *The American Economic Review* 61, no. 4 (1971): 561-574.

27) Holmstrom, Bengt. “Understanding the Role of Debt in the Financial System.” BIS Working Papers No. 479. Bank for International Settlements, 2015.

28) Holmstrom, Bengt, and Jean Tirole. "Financial intermediation, loanable funds, and the real sector." *the Quarterly Journal of Economics* 112, no. 3 (1997): 663-691.

29) Kaufman, George G. "Too big to fail in banking: What does it mean?." *Journal of Financial Stability* 13 (2014): 214-223.

30) Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. "The aggregate demand for treasury debt." *Journal of Political Economy* 120, no. 2 (2012): 233-267.

31) Liikanen, Erkki. "High-level Expert Group on reforming the structure of the EU banking sector." *Final Report, Brussels 2* (2012).

32) Merton, Robert C. "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory." *Journal of Banking & Finance* 1, no. 1 (1977): 3-11. Chapter 19 in Continuous-Time Finance, Blackwell: 1992b.

33) Merton, Robert C. "On the Application of the Continuous-time Theory of Finance to Financial Intermediation and Insurance." *Geneva Papers on Risk and Insurance* (1989): 225-261.

34) Merton, Robert C. "The financial system and economic performance." In *International Competitiveness in Financial Services*, pp. 5-42. Springer Netherlands, 1990.

35) Merton, Robert C. "Financial Intermediation in the Continuous-Time Model", in *Continuous-Time Finance*, Chapter 14, 428-471. Blackwell: 1992a.

36) Merton, Robert C. "On the Cost of Deposit Insurance When There are Surveillance Costs", *Journal of Business*, 51 (July 1978), pp 439-52. In *Continuous-Time Finance*, Chapter 20. Blackwell: 1992b.

37) Merton, Robert C., "Operation and Regulation in Financial Intermediation: A Functional Perspective", in *Operation and Regulation of Financial Markets*, ed. Peter Englund, (The Economic Council, Stockholm, 1993), 17-67.

38) Merton, Robert C. "A Functional Perspective of Financial Intermediation." *Financial Management* 24, no. 2 (1995): 23-41.

39) Merton, Robert C. "A Model of Contract Guarantees for Credit-Sensitive, Opaque Financial Intermediaries." *European Finance Review* 1, no. 1 (1997): 1-13.

40) Merton, Robert C., and Zvi Bodie. "A Conceptual Framework for Analyzing the Financial Environment." Chap. 1 in *The Global Financial System: A Functional Perspective*, by D. B. Crane et. al., 3–31. Boston: Harvard Business School Press, 1995.

41) Merton, Robert C., and Zvi Bodie. "Design of Financial Systems: Towards a Syntheses of Function and Structure." *Journal of Investment Management* 3, no. 1 (2005): 6.

42) Myers, Stewart C., and Nicholas S. Majluf. "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics* 13, no. 2 (1984): 187-221.

43) Ramakrishnan, Ram TS, and Anjan V. Thakor. "Information reliability and a theory of financial intermediation." *The Review of Economic Studies* 51, no. 3 (1984): 415-432.

44) Stiglitz, Joseph E., and Andrew Weiss. "Credit rationing in markets with imperfect information." *The American Economic Review* 71, no. 3 (1981): 393-410.

45) Vickers, John Sir. *Independent Commission on Banking final report: recommendations*. The Stationery Office, 2011.

46) Wakker, Peter, Richard Thaler, and Amos Tversky. "Probabilistic Insurance." *Journal of Risk and Uncertainty* 15, no. 1 (1997): 7-28.

47) Zenios, Stavros A. "Fairness and reflexivity in the Cyprus bail-in." *Empirica* 43, no. 3 (2016): 579-606.

Appendix

Proof of Theorem 1: Note that $\frac{\partial ES_{total}}{\partial p} = \frac{\partial ES_c}{\partial p} = V$. Moreover, $\frac{\partial^2 ES_{total}}{\partial p \partial V} = \frac{\partial^2 ES_c}{\partial p \partial V} > 0$. Now p^0 is the value of p that satisfies $ES_c = 0$, so $p^0 = \overline{V}/V$. Similarly, p^* is the value of p that satisfies $ES_{total} = 0$. Thus, $p^* = \left[\overline{V} + k - V_m^I\right]/V$. This yields:

$$1 - p^0 = [V - \overline{V}][V]^{-1} \tag{A-1}$$

$$1 - p^* = [V - \overline{V} + V_m^I - k][V]^{-1} \tag{A-2}$$

It follows from (1) and (2) that $1 - p^0 \geq 0$ and $1 - p^* > 1 - p^0$. Finally, note that $\frac{\partial p^*}{\partial [V_m^I - k]} = -\frac{1}{V} < 0$.

■

Proof of Theorem 2: The net expected surplus to the customer is

$$ES_c = [\xi u_l + \{1 - \xi\}F + u_s]p - f \tag{A-3}$$

and the surplus to the intermediary is

$$f - pF - k \tag{A-4}$$

so the total expected net economic surplus is:

$$ES_{total} = pu_s + \xi p[u_l - F] - k \tag{A-5}$$

which is clearly maximized at $p = 1$. Moreover, the assumption that intermediation has social value guarantees that $u_s + \xi[u_l - F] > k$. ■

Proof of Corollary 1: With perfect competition among intermediaries, we have

$$f(p_1) = p_1 F - k \tag{A-6}$$

$$f(p_2) = p_2 F - k \tag{A-7}$$

so the customer pays a lower price to the intermediary with solvency probability p_2 than to the one with solvency probability p_1, i.e., $f(p_2) < f(p_1)$. However, using (A-6) and (A-7):

$$ES_c(p_1) = \xi p_1[u_l - F] + p_1 u_s - k \tag{A-8}$$

$$ES_c(p_2) = \xi p_2[u_l - F] + p_2 u_s - k \tag{A-9}$$

Clearly, $ES_c(p_1) > ES_c(p_2)$. ■

Proof of Corollary 2: Using the proofs of Theorem 2 and Corollary 1, we see that if f is set to

enable the intermediary to fully extract all customer surplus, then $f = p\xi[u_l - F] + pF + pu_s$, which means the surplus for the intermediary is $p\xi[u_l - F] + pu_s - k$, which is strictly increasing in p. ■

Proof of Corollary 3: The repayment on the borrowing is $[1 - p][F + u_s]$. If the intermediary does not fail, the customer receives $F + u_s$. After repaying the loan used to purchase the guarantee, the customer is left with

$$F + u_s - [1 - p][F + u_s]$$
$$= p[F + u_s] \quad \text{(A-10)}$$

The total expected net economic surplus associated with this is

$$pu_s + p\xi[u_l - F] - k \quad \text{(A-11)}$$

If the intermediary fails, the customer receives $F + u_s$ from the guarantor, so the customer is again left with the expression in (A-10) after repaying the loan, and a total expected net economic surplus that is the same as in (A-11). So the overall total expected net economic surplus is that in (A-11), which is exactly ES_{total} in (A-5). ■

Proof of Lemma 1: The first-best choice of e maximizes:

$$p(e)[X + V_s] - \varphi(e) \quad \text{(A-12)}$$

The first-order condition is:

$$p'(e^0)[X + V_s] - \varphi'(e^0) = 0 \quad \text{(A-13)}$$

The second-order condition (which clearly holds here) is:

$$p''(e^0)[X + V_s] - \varphi''(e^0) < 0 \quad \text{(A-14)}$$

The result that $e^0 \in (0,1)$ follows from the Inada conditions on φ.

Next, substituting (9) and (8) into (6), we can express it as:

$$r^{-1}\{p(e)[X + V_s] + [1 - p(e)]I_0 V_s + [1 - p(e)]I_1\lambda_0 V_s - fr - k - \varphi(e)\} \quad \text{(A-15)}$$

It is clear that, setting $e = e^0$ and choosing $I_0 = 1$ maximizes (A-15). Since in the first best, we can ignore (7), $\gamma = 1$ is optimal. ■

Proof of Theorem 3: In the second best, if $\gamma = 1$, (7) becomes

$$e \in \underset{e \in [0,1]}{\operatorname{argmax}}\, r^{-1}\{x - F - \varphi(e)\} \quad \text{(A-16)}$$

Clearly, $e^* = 0$ is the solution. Thus, $\gamma = 1$ cannot be optimal. The reason is that $p(0) = 0$, so by (9),

$gr = X$, which means that the bank's objective function (6) becomes (after substituting for g), $r^{-1}\{-F - k\} < 0$. The result that $\gamma = \gamma_0$ follows from the fact that this preserves $\lambda_0 V_s$ of the depositor service value, and the cost of purchasing the guarantee is a net wash against the offsetting adjustment in the repayment obligation to depositors. This clearly dominates setting $\gamma < \gamma_0$ and losing λV_s. ■

Proof of Lemma 2: The first-order condition on e corresponding to (12) is:

$$[\alpha fx + [1 - \alpha]fzq - F][1 - q] - \varphi' = 0 \tag{A-17}$$

after substituting $p(e) = e[1 - q]$. The second-order condition is:

$$-\varphi'' < 0 \tag{A-18}$$

which clearly holds. Now using (13) and totally differentiating the first-order condition (A-17), we get:

$$f[x - qz][1 - q] + [1 - q]^2 V_s \frac{de^*}{d\alpha} - \varphi'' \frac{de^*}{d\alpha} = 0 \tag{A-19}$$

So

$$\frac{de^*}{d\alpha} = \frac{f[1 - q][x - qz]}{\varphi'' - [1 - q]^2 V_s} > 0 \tag{A-20}$$

Thus,

$$\frac{d^2 e^*}{d\alpha^2} = \frac{-f[1 - q][x - qz]\varphi'''}{[\varphi'' - [1 - q]^2 V_s]^2}$$

$$< 0 \tag{A-21}$$

since $\varphi''' > 0$. ■

Proof of Theorem 4: Using (13), we have

$$F = fr - [q + p(e)]V_s \tag{A-22}$$

Thus,

$$\frac{\partial F}{\partial \alpha} = -[1 - q]V_s \frac{de^*}{d\alpha} < 0 \tag{A-23}$$

and

$$\frac{\partial^2 F}{\partial \alpha^2} = -[1 - q]V_s \frac{d^2 e^*}{d\alpha^2} > 0 \tag{A-24}$$

Now, substituting for F from (A-22) into (11), and using the Envelope Theorem, we can write the first-order condition for α as:

$$fx[q + [1 - q]e] - fqz - \frac{\partial F}{\partial \alpha} = 0 \tag{A-25}$$

and the second-order condition is:

$$-\frac{\partial^2 F}{\partial \alpha^2} < 0 \tag{A-26}$$

which clearly holds since $\frac{\partial^2 F}{\partial \alpha^2} > 0$. Now totally differentiate the first-order condition (A-25) to get:

$$-fq - \frac{\partial^2 F}{\partial \alpha^2}\left[\frac{d\alpha^*}{dz}\right] = 0 \tag{A-27}$$

So,

$$\frac{d\alpha^*}{dz} = -\frac{fq}{\partial^2 F / \partial \alpha^2} < 0 \tag{A-28}$$

Substituting (A-23) into (A-25), we have:

$$fx[q + [1 - q]e] - fqz + [1 - q]V_s \frac{de^*}{d\alpha} = 0 \tag{A-29}$$

Note that (A-20) and (A-29) imply that in equilibrium it must be true that:

$$qz > x[q + [1 - q]e^*] \tag{A-30}$$

In other words, the choice of α^* is always such that market-based activities have a higher expected payoff per dollar invested. ■