Review of Finance, 2016, 2219–2246
doi: 10.1093/rof/rfv069
Advance Access Publication Date: 27 January 2016

Banking and Trading*

Arnoud W. A. Boot[1] and Lev Ratnovski[2]

[1]University of Amsterdam and CEPR and [2]International Monetary Fund

Abstract

We study the interaction between relationship banking and short-term arm's length activities of banks, called trading. We show that a bank can use the franchise value of its relationships to expand the scale of trading, but may allocate too much capital to trading *ex post*, compromising its ability to build relationships *ex ante*. This effect is reinforced when trading is used for risk shifting. Overall, combining relationship banking and trading offers benefits under small-scale trading, but distortions may dominate when trading is unbridled. This suggests that trading by banks, while benign historically, might be distortive with deeper financial markets.

JEL classification: G21, G24, G28, G32

1. Introduction

The recent crisis has reopened the debate on the merits of combining traditional commercial banking with market-based activities. This article sheds light on a novel interaction between relationship banking and short-term arm's length activities of banks, which we call "trading." We focus on differences in the time horizon and scalability of relationship banking and trading, and study the desirability of combining them. We show that although some trading by banks may be profitable, the deepening of financial markets might lead to time-inconsistency problems in capital allocation, where banks engage in too much trading at the expense of relationship banking, to the detriment of shareholder value.

Relationship banking involves private information and repeated interactions with an established set of customers. This makes relationship banking a long-term and not easily

* We thank Franklin Allen (the Editor), an anonymous referee, Tobias Adrian, Stijn Claessens, Giovanni Dell'Ariccia, Giovanni Favara, Thomas Gehrig, Hans Gersbach, Linda Goldberg, Itay Goldstein, Mark Flannery, Jan-Pieter Krahnen, Luc Laeven, Christian Laux, Richhild Moessner, Marco Pagano, George Pennacchi, Joel Shapiro, Kostas Tsatsaronis, and participants of the ETH-NYU Law Conference on "Tackling Systemic Risk", Basel Committee-CEPR Conference on "Banks—How Big is Big Enough", Toulouse Conference "Risk Management and Financial Markets", Oxford Financial Intermediation Conference, FDIC Annual Conference "Financial Services in the Current Environment", Swiss Winter Conference on Financial Intermediation, and New York Fed-LBS conference on "Global Banks," the Bonn conference on "The Structure of Banking Systems," and of seminars at IMF and BIS for helpful comments. The views expressed are those of the authors and do not necessarily represent those of the IMF.

scalable activity. Trading does not rely on private information. When markets are deep enough, a bank can go in and out of positions rapidly. This makes trading relatively short term and scalable. In modern banks, trading covers activities such as proprietary trading, investing in asset-backed securities, carry trade, market-making, originating loans based solely on hard information (typically mortgages), and all other activities that do not rely on relationship-specific private information. Observe that what we call trading could include what the literature has called "transaction lending" (e.g., mortgage origination based on hard information) as this does not rely on relationship-specific private information, is typically scalable, and leads readily to tradeable claims.

The key contribution of this article is to show that banks might be naturally induced to expand *short-term* trading-like activities too much, undermining their *long-term* relationship banking franchise. The main channel for that in our analysis is a time-inconsistency problem in bank capital allocation. We focus on the arrangement where relationship banks provide their customers with funding insurance (i.e., guarantees for the availability of funding) via credit lines or via informal commitments to local markets. In the case of credit lines, the upfront fees paid for funding insurance lead to "front-loaded" (received in advance) income for banks. Moral hazard at the borrower level prevents banks from offering fully committed funding insurance; banks maintain discretion on whether to lend. Combined, discretion, and upfront fees invite time-inconsistency problems. In particular, banks may opportunistically shift capital to trading *ex post*, once the upfront funding insurance fees have been collected, and in doing so reduce their ability to honor the funding commitments. Anticipating this, customers will be unwilling to build relationships with banks and the relationship franchises may suffer. In the case of a bank's commitment to local markets a similar mechanism is at play. Banks that can commit to standing with their borrowers in difficult times can charge higher *ex ante* prices. Again, time inconsistency may undermine these commitments. Key is that access to trading opportunities may undermine a bank's ability to serve the more long-term oriented relationship banking business. In our formal model, this runs via not being able to honor future funding commitments, but we see it as a more general point caused by the difference in horizon between relationship banking and trading. Trading, including more scalable and possibly more opportunistic transactional lending, is more short-term oriented and dependent on market momentum (e.g., the booming 90 s), and may then gain the upper hand at the expense of the more "boring" steady relationship banking business.[1]

The market failure highlighted by our analysis is consistent with multiple strands of the empirical literature. First, the funding insurance role of relationship banks, while sometimes overlooked in the literature, is empirically very important. Bank provides funding insurance through credit lines (Kashyap, Rajan, and Stein, 2002) and—importantly—through implicit commitments to customers in local markets, particularly during crises (Bolton *et al.*, 2013; Sette and Gobbi, 2015).[2] Second, banks possess and use discretion as to whether to honor lending commitments. Although outright breaches of lending commitments are rare, banks

1 Note from Adrian and Shin (2013) that trading-like bank activities, such as investment banking, are cyclical and expand substantially during upturns, while traditional, commercial banking is less cyclical.

2 Bae, Kang, and Lim (2002) show that banks in Korea offered their relationship customers better credit terms than those available in the spot market during the 1997–98 crisis. Beck *et al.* (2013) show that relationship banks constrained credit less than transactions-oriented banks in Eastern Europe during the 2008 credit crisis. Petersen and Rajan (1994) originated this strand of empirical literature by showing that banks offer their relationship customers better availability of credit.

use overly tight covenants and (to a lesser extent) material adverse change (MAC) clauses to renege on credit lines (Chava and Roberts, 2008; Sufi, 2009; Demiroglu and James, 2011). The literature confirms that banks renege on funding commitments to customers when their balance sheet strength is compromised (Ivashina and Scharfstein, 2010; Acharya *et al.*, 2014a, 2014b). Trading is a common factor behind bank balance sheet weakness.[3] Interestingly, and consistent with our analysis, banks' commitment to relationship customers can also be compromised when trading opportunities are strong. Chakraborty, Goldstein, and MacKinlay 2014) show that banks with superior opportunities in transactional mortgage lending reduce their lending to bank-dependent commercial borrowers, forcing them to cut investment.

Overall, the empirical literature confirms, perhaps strikingly, that trading activities may reduce the value of banks. Laeven and Levine (2007) show that banks that combine lending and non-lending activities lose value compared with engaging in these activities separately. Similar evidence is offered by Stiroh (2004), Mercieca, Schaeck, and Wolfe (2007), Schmid and Walter (2009), and Baele, De Jonghe, and Vander Vennet (2007). Our model is the first one to explain this dynamics. We show that banks may inefficiently divert resources to trading as a result of a time-inconsistency problem, to the detriment of relationships-oriented activities and overall bank profitability.

The outline of the analysis is as follows. We start by describing the synergy between relationship banking and trading. Relationship banking has information-based rents that generate implicit capital, yet is not readily scalable. The trading activity is scalable but has lower margins, so can be capital constrained. Accordingly, relationship banks may expand into trading in order to profitably use their "spare" capital.[4] Opening up banks to trading, however, creates distortions. The main distortion is a time-inconsistency problem in capital allocation, which leads to excessive trading. A complementary distortion is the risk shifting that may be facilitated by trading. We show that this distortion reinforces the time-inconsistency problem. The distortions intensify when financial markets are deeper, allowing larger trading positions, and when returns on relationship banking are lower. Both trends—deeper financial markets and less profitable relationship banking—were present in the last decades.[5] This could explain

3 De Jonghe (2010), Demirguç-Kunt and Huizinga (2010), Brunnermeier, Dong, and Palia (2012), and DeYoung and Torna (2013) show that banks' non-lending activities are riskier than lending. De Jonghe (2010) and Brunnermeier, Dong, and Palia (2012) show that within the non-lending activities trading is the riskiest. Fahlenbrach, Prilmeier, and Stiltz (2012) show that banks with more exposure to trading securities had higher losses during the 1998 and 2008 crises.

4 This liability-side synergy is akin to the assertions of practitioners that one can "take advantage of the balance sheet" of the bank. Indicators of the shift of US banks toward trading include an increase in the ratio of trading assets and securities to loans on bank balance sheets from 30% in early 1990s to 60% in 2013, and an increase of non-interest income as a share of bank revenues from 35% to 50% over the same period (NY Fed, 2014). The Liikanen Report (2012) points at similar developments in Europe.

5 The net interest margin of US banks has been steadily declining from over 4% in mid-1990s to 3–3.5% in early 2000s and just above 2.5% in 2012–13 (NY Fed, 2014). See also Edwards and Mishkin (1995) and DeYoung and Rice (2004) for a discussion of declining profitability in relationship banking. The deepening of financial markets shows up, for example, in the ratio of US stock market capitalization to GDP that has tripled and the ratio of US private bond market capitalization to GDP which almost doubled from the early 1990s to 2007, with a minor pullback after the crisis (Beck, Demirgüç-Kunt, and Levine, 2010).

why trading by banks, while benign historically, might recently have become distortive from the perspective of its impact on the bank's relationship franchise.

The notion that trading activities can undermine banks goes back at least to the UK Barings Bank trading disaster in Singapore in 1995. But the concerns were especially vivid during the recent financial crisis. Many European universal banks suffered dramatic losses on their asset-backed securities holdings or exposures to sponsored investment vehicles (UBS in Switzerland is a good example; UBS, 2008). Similarly in the USA, investments in asset-backed securities have backfired in both investment banks (Bear Stearns, Lehman Brothers and Merrill Lynch) and commercial banks (Washington Mutual and Wachovia). This article highlights that, besides outright losses, trading may undermine banks also in a more subtle way: by diverting resources away from the relationship banking activity.

The interaction between banks and financial markets is a rich research area. Much of the earlier literature focused on the interaction between lending and underwriting, and its implications for the merits of the Glass-Steagall Act in the USA and universal banking in Europe (Kroszner and Rajan, 1994; Puri, 1996; Schenone, 2004; Fang, Ivashina, and Lerner, 2010). A key question in that literature is whether private information induces positive or negative synergies between lending and underwriting. We instead contrast relationship banking to an activity that does not rely on private information—trading. There is no scope for informational spillovers, so we identify different relevant interactions. Somewhat closer to our analysis are papers that study how the expansion of financial markets affects the nature of banking—by impacting the depth of relationships (Boot and Thakor, 2000; Hauswald and Marquez, 2006) or making banking more cyclical (Shleifer and Vishny, 2010). Our article focuses on the time-inconsistency problem in bank capital allocation, driven by differences in the time horizon and scalability of relationship banking and trading. The focus on time-inconsistency differentiates this paper from related work by Kahn and Winton (2004) who study risk spillovers between bank activities, and Boyd, Chang, and Smith (1998) and Freixas, Lóránth, and Morrison (2007) who focus on the abuse of the safety net.[6] Our analysis also relates to the literature on the structure and scope of financial institutions. We highlight that the deepening of financial markets may have induced diseconomies of scope in combining trading and relationship banking.

The article is organized as follows. Section 2 sets up the model. Section 3 shows the synergy between relationship banking and trading. Section 4 analyzes the time-inconsistency problem in capital allocation, including its interaction with risk shifting. Section 5 discusses robustness issues and empirical and policy implications. Section 6 concludes.

2. Model

We model two activities. One is relationship banking (banking); the other is short-term arm's length operations (trading). Banking relies on an endowment of private information

6 More generally, our paper relates to the literature on internal capital markets (Williamson, 1975; Donaldson, 1984). Conglomeration may relax the firm's overall credit constraint (Stein, 1997) but can lead to distortions (Rajan, Servaes, and Zingales, 2000). Our model considers a specific distortion: a misallocation of capital due to the time-inconsistency problem in combining a long-term commitment-based business with short-term opportunities that may compromise commitment. Boot, Greenbaum, and Thakor (1993) show that commitment problems in banks could be alleviated by reputation building.

about an established set of customers; trading does not. We argue that this observation alone suffices to highlight a range of distinctions between the two. The information endowment makes banking profitable (hence, we assume, not credit constrained), yet not easily scalable. Securing the value of information requires multiple interactions with customers, so banking is long term in nature. In contrast, since it does not rely on an information endowment, trading is scalable, less profitable per unit (and hence, we assume, credit constrained), and short-term oriented. We use these distinctions to study synergies and conflicts between banking and trading.

2.1 Credit Constraints

Before describing the banking and trading activities, we introduce a key modeling feature: the presence of credit constraints. We build on Holmstrom and Tirole's (1998) formulation where a firm's borrowing capacity is limited to a multiple of the owner-manager's net worth. Assume that the owner manager can run the bank normally or "shirk". Shirking makes bank assets worthless, but gives the owner-manager private benefits proportional to firm size. She will run the bank normally when:

$$\Pi \geq bA, \tag{1}$$

where Π is the shareholder return when assets are employed for normal bank operations (in the model, the owner-manager has no other wealth) and bA is return to shirking: the initial investment A multiplied by the conversion factor b, $0 < b < 1$, of assets into private benefits. Returns to shirking can be interpreted as the consumption of perquisites when not working, or as the payoff to absconding in the "take the money and run" approach, cf. Calomiris and Kahn (1991). Anticipating the need to ensure the owner-manager's effort, bank creditors will only provide funding for assets A insofar as the constraint (1) is satisfied, making it a borrowing (leverage) constraint for a bank.

2.2 Banking

We model relationship banking as a profitable, yet not scalable business, which therefore is not credit constrained. A bank operates in a risk-neutral economy with no discounting. It has no explicit equity and has to borrow in order to lend. Bank creditors require a zero expected return. In the base model all activities are risk-free, so the interest rate on bank borrowing is zero. There are three dates: 0, 1, and 2.

At date 0 the bank is endowed with private information on a mass $\overline{R}$ of customers. The bank can earn rents in two ways:

- First, the bank has a fixed profit R_0, coming from services that are largely invariant to the rest of a bank's business; payment services might be an example of this.
- Second, the bank has an opportunity to serve its customers' funding needs. Each customer needs to borrow 1 unit at date 1, to repay at date 2 (with certainty). When a bank lends, the information endowment enables it to collect informational rents r per customer. The total rents are then rR, where $R \leq \overline{R}$ is the volume of relationship banking activity—the amount that the bank borrows from the market and lends to its customers.

The bank collects profits R_0 and rR (the latter in the form of higher interest rates) at date 2. Under this setup, the leverage constraint (1) for the bank takes the form:

$$R_0 + rR \geq bR, \tag{2}$$

where the left-hand side is the bank's profit in normal operations and the right-hand side is the moral hazard payoff. We assume that this constraint is satisfied, including at $R = \overline{R}$, implying spare borrowing capacity in banking:

$$R_0 + r\overline{R} > b\overline{R}. \tag{3}$$

Note two assumptions implicit in this characterization of relationship banking. The first assumption is that the information endowment is fixed, so that relationship banking is not easily scalable. Indeed, expanding the relationship banking customer base may be costly because of adverse selection (Dell'Ariccia and Marquez, 2006) or difficulties in processing large amounts of soft information (Gale, 1993; Stein, 2002). The second assumption is that a relationship bank has some informational monopoly over its borrowers, enabling it to earn rents. This might be related to past investments by the bank and its customers in their relationships, and/or to the advantages of proximity or specialization in local markets. The time and proximity elements involved in building relationships also provide an additional explanation for the lack of scalability.

2.3 Trading

We model trading as a scalable but less profitable business, which consequently is credit constrained. For T units invested at date 1, trading produces at date 2 net returns tT for $T \leq S$ and 0 for each unit exceeding S, thus tS for $T > S$. Trading is less profitable per unit than banking since it does not benefit from the informational endowment[7]:

$$t < r. \tag{4}$$

And the low profitability of trading makes it credit constrained:

$$t < b, \tag{5}$$

implying that the leverage constraint (1) does not hold when trading is a standalone activity. This implies that, in this model, standalone trading is impossible, despite the opportunity to profitably invest up to S units. This simplifying assumption should not be taken literally. It only means that—consistent with practice—most trading operations require a substantial equity commitment.[8]

7 Trading by banks is indeed less profitable per unit than lending. In 2000–2007, the average return on banks' trading assets was 2% (during the crisis, the return was negative). In the same period, the average bank net interest margin was 3.25% (NY Fed, 2014) and the average cost of bank funding 3% (according to the Federal Home Loan Bank of San Francisco Cost of Funds Index), making the gross return on lending 6%. Another useful observation is that most trading by banks is volume-based (e.g., carry trade); it is different from—and has a lower return—than proprietary strategy-based trading by hedge funds. Overall, one could characterize relationship banking as a high-margin-low-volume operation and trading by banks as a low-margin-high-volume operation.

8 For example, hedge funds (or other independent trading houses) are normally partnerships with equity commitments by the partners that facilitate substantial recourse. Note that we do not consider external equity. A lack of access to external equity can be rationalized in a standard way

Date 0

- Bank is endowed with a customer base $\overline{R}$.

Date 1

- Bank customers have liquidity needs;
- Bank chooses the allocation of borrowing capacity between relationship banking (covering customers' liquidity needs) $\boldsymbol{R}$ (to earn $\boldsymbol{r}$) and trading $\boldsymbol{T}$ (to earn $\boldsymbol{t}$);
- Bank borrows and can divert resources to moral hazard, which would obtain a payoff $\boldsymbol{b(R+T)}$;
- Bank lends to customers and engages in trading.

Date 2

- Returns are realized, everyone is repaid.

Figure 1. The timeline.

The parameter S captures the scalability of trading. In a richer model, one could think of decreasing returns to scale in the context of a Kyle (1985) framework, where the average return of an informed trader falls in the size of her trade, but less so when the mass of liquidity traders is larger. Thus, S is higher—trading is more scalable—in deeper financial markets, or for higher financial development.

The timeline for the benchmark model is summarized in Figure 1. The bank maximizes shareholder wealth and prefers banking to trading when indifferent.[9]

3. Synergy between Banking and Trading

Our model implies a natural benefit to combining banking and trading: it links a business that has borrowing capacity but lacks scalability (the relationship bank) with a business that has investment opportunities but is subject to credit constraints (trading).

Under conglomeration, at date 1, the bank maximizes its profit:

$$\Pi_C = R_0 + rR + tT, \tag{6}$$

subject to the joint leverage constraint (use Equation (1)):

$$R_0 + rR + tT \geq b(R + T), \tag{7}$$

where $T \leq S$. Comparing this to Equations (3) and (5) shows that conglomeration allows using the borrowing capacity of the relationship bank to fund some trading.

In allocating the borrowing capacity between relationship banking R and trading T the bank chooses to serve all banking customers first ($R = \overline{R}$) before allocating any funds to trading, because banking is more profitable: $r > t$. The maximum amount of trading that a

through agency costs. In Section 5.4, we consider how the credit constraints problem in trading can be affected by the ability to pledge trading assets.

9 An alternative to shareholder value maximization would be to focus on the incentive problems of managers (e.g., as in Acharya, Pagano, and Volpin, 2013). While managerial incentive issues in banks are undoubtedly important, understanding the distortions that can be caused by shareholder value maximization alone remains critical, particularly when such distortions may have worsened recently due to external factors, such as information technology-linked increases in financial development.

bank can support, $T_{\max}$ (assuming $T_{\max} \leq S$), is given by Equation (7) set to equality, with $R = \overline{R}$:

$$T_{\max} = \frac{R_0 + \overline{R}(r - b)}{b - t}. \tag{8}$$

Since it is never optimal to trade at a scale that exceeds S, $T = \min\{S, T_{\max}\}$. This means that when the scalability of trading is small—S is low—the bank covers all profitable opportunities in trading. When trading is more scalable—S is high—the bank covers trading opportunities $T_{\max} < S$ and abstains from the rest.

Proposition 1 (Synergy between banking and trading). *The relationship bank can use its implicit capital to expand the scale of credit-constrained trading. In equilibrium, the bank serves all relationship banking customers, $R = \overline{R}$, and allocates the rest of its borrowing capacity to trading, as long as trading is profitable: $T = \min\{S, T_{\max}\}$.*

Proposition 1 is a benchmark that explains why banks may choose to engage in trading and specifies the first-best allocation of borrowing capacity between the two activities. Next, we study inefficiencies that may arise in combining banking and trading.

4. Time Inconsistency of Bank Capital Allocation

The previous section has outlined the synergy between banking and trading—the use of "spare" capital of the relationship bank to expand trading. We now turn to the cost of conglomeration: the time-inconsistency problem in bank capital allocation, driven by a conflict between the long-term nature of relationship banking and short-term trading.

Relationship banking is long term because it involves repeated interactions with customers, with returns distributed over time. We model this intertemporal nature by letting a bank offer funding commitments to customers. The commitments take the form of credit lines that cover future funding needs in return for *ex ante* fees that customers pay to the bank. A rationale for credit lines is that borrowers may be constrained in how much they can pay to the bank *ex post*, due to moral hazard at that stage (Holmstrom and Tirole, 1998; Bolton *et al.*, 2013; Acharya *et al.*, 2014a). The more favorable *ex post* lending terms under the commitment with *ex ante* fees may contain that moral hazard.

However at the level of the bank, the *ex ante* fees may cause time-inconsistency problems. By making some payments *ex ante*, customers reduce the amount that they owe the bank *ex post*. As a result, returns to banking, although higher than returns to trading overall, might *ex post* be lower. This may distort capital allocation: once credit line fees have been collected, a bank might be induced to allocate capital primarily to trading, leaving itself with insufficient borrowing capacity to fully serve the credit lines. Anticipating that, customers would reduce the credit line fees that they are willing to pay *ex ante*, lowering the bank's overall profit and borrowing capacity.

A salient feature of credit line contracts is that the lender has discretion on whether to honor lending commitments. The main analysis takes a simplified credit line contract as given, including the lender's discretion to honor such contract, and analyzes its consequences. In Section 5.3, we show that such a contract arises as an optimal response to two distortions: limited pledgeability of assets, which restricts the borrower's ability to repay the bank in some states of the world, and moral hazard at the borrower level, which prevents fully committed (i.e., non-discretionary) funding arrangements.

The emphasis on the "funding insurance" role of relationship banks is a key feature of our model. This approach captures not only credit lines, but also banks' implicit commitments to customers in local markets. As we will argue in Section 5.1, we see such commitments as a general feature of relationship banking, which contrasts with the informational capture view that has been commonly emphasized in the literature.

4.1 Setup

Assume that while a bank can generate a return r on covering its customers' future funding needs, it can only capture $\rho \leq r$ through interest rates charged on the actual lending between dates 1 and 2. We treat the maximum *ex post* charge on bank lending ρ as an exogenous parameter, but it represents the maximum repayment that a bank can charge without inducing moral hazard at the borrower level (see Section 5.3). The remaining $(r-\rho)$ can be captured at date 0 as a credit line fee. Importantly, a bank cannot commit to cover the future liquidity needs of customers. That is, we let the bank have discretion to refuse lending in the future if it has no borrowing capacity left to lend under the credit line.[10] The timeline incorporating the credit line arrangement is shown in Figure 2.

At date 1, the bank chooses R and T to maximize its profit:

$$\Pi_C = R_0 + (r-\rho)R_{\text{anticipated}} + \rho R + tT, \tag{9}$$

where $(r-\rho)R_{\text{anticipated}}$ is the total credit line fee received by the bank at date 0, $R_{\text{anticipated}}$ is the volume of borrowing that customers expect to get under the credit line, and R is the actual borrowing under the credit line. The bank maximizes Equation (9) subject to the leverage constraint:

$$R_0 + (r-\rho)R_{\text{anticipated}} + \rho R + tT \geq b(R+T). \tag{10}$$

Note that in equilibrium, $R_{\text{anticipated}} = R$. This follows because customers correctly anticipate the bank's ability and willingness to lend under the commitment.

Recall that in the benchmark case in Section 3, the bank always covers the customers' liquidity needs first, because the return on the banking activity exceeds that on trading $(r > t)$. However, when part of the return to banking is obtained *ex ante* as credit line fees, a time-inconsistency problem may distort capital allocation. As long as the *ex post* return to banking is sufficiently high, $\rho \geq t$, the first best allocation persists. Yet, when $\rho < t$, the bank chooses to allocate the borrowing capacity to trading up to its maximum profitable scale S first, and only then to give the remainder to banking. This constitutes a time-inconsistency problem in bank capital allocation. In what follows, we analyze its consequences.[11]

10 Here, for simplicity, we set the probability that the credit line will be used equal to one. In practice, the probability is usually less than one: the firm only draws on the credit line in states where external circumstances make spot markets too expensive. Also, in a richer model, the actual use of a credit line may coincide with a more stressful environment for banks. In such setting, the possibility of having to renege on commitments may come about even in absence of time-inconsistency problems. In Boot, Greenbaum, and Thakor (1993), this is used to rationalize the MAC clause.

11 As Section 5.3 shows, the bank will always charge the maximum feasible *ex post* payment ρ (maximum that does not trigger moral hazard at the borrower level) to minimize the time-inconsistency problem. Note also that in a multi-period setting the future value of the relationship business could

Date 0

- Bank is endowed with a customer base $\overline{R}$;
- Bank collects credit line fees $\boldsymbol{(r-\rho)R_{\text{anticipated}}}$.

Date 1

- Bank customers have liquidity needs;
- Bank chooses the allocation of borrowing capacity between relationship banking (covering customers' liquidity needs) $\boldsymbol{R}$ (to earn $\boldsymbol{\rho}$) and trading $\boldsymbol{T}$ (to earn $\boldsymbol{t}$);
- Bank borrows and can divert resources to obtain a moral hazard payoff $\boldsymbol{b(R+T)}$;
- Bank lends to customers and engages in trading.

Date 2

- Returns are realized, everyone is repaid.

Figure 2. The timeline with a credit line arrangement.

The severity of the time inconsistency depends on the scalability of trading and the returns on relationship banking. When the scalability of trading S is low and/or the profitability of relationship banking r is high (making $T_{\max}$ high, see Equation (8)), so that $S \le T_{\max}$, the bank can cover all liquidity needs of customers $\overline{R}$ even after allocating S to trading. Then, trading does not trigger time inconsistency. However, when the scalability of trading S is high and/or the return to relationship banking r is low, so that $S > T_{\max}$, the relationship banking activity is credit constrained *ex post*. The borrowing capacity that remains after the bank has allocated S to trading is insufficient to cover all the funding needs of customers. Relationship banking customers correctly expect $R_{\text{anticipated}} < \overline{R}$, so they pro-rate the credit line fee that they are willing to pay *ex ante*: $(r-\rho)R_{\text{anticipated}} < (r-\rho)\overline{R}$. The profits of the relationship bank on lending to customers fall from $r\overline{R}$ to $rR_{\text{anticipated}}$. As we will show, this affects the scale at which the bank can invest and hence the amount of banking and trading activity that can be undertaken because it lowers the bank's borrowing capacity.

We can now calculate the equilibrium capital allocation. Cutting down on relationship banking R to accommodate more trading conserves capital at a rate b but undermines bank profitability at a rate r (see the leverage constraint (1)). When r is small, such that $r < b$, cutting down on banking frees up more capital than is lost in profits (the leverage constraint becomes slack). Hence, as S starts exceeding $T_{\max}$, trading expands while banking contracts smoothly. We solve Equation (10) set as equality to get the equilibrium allocation (note from Equation (10) that $T_{\max} = T_{\max}(r)$):

$$R = \begin{cases} \overline{R}, \text{ for } S \le T_{\max} \\ \dfrac{R_0 - S(b-t)}{b-r}, \text{ for } T_{\max} < S \le \dfrac{R_0}{b-t} \\ 0, \text{ for } S > \dfrac{R_0}{b-t} \end{cases}, \quad T = \begin{cases} S, \text{ for } 0 < S \le \dfrac{R_0}{b-t} \\ \dfrac{R_0}{b-t}, \text{ for } S > \dfrac{R_0}{b-t} \end{cases}. \tag{11}$$

mitigate the time-inconsistency problem in credit lines. In the model this could be captured by $\rho^* > \rho$ that incorporates future profits. Our results are therefore more likely to hold when future profits are insufficient to prevent opportunistic behavior (Keeley, 1990), that is, ρ^* is small.

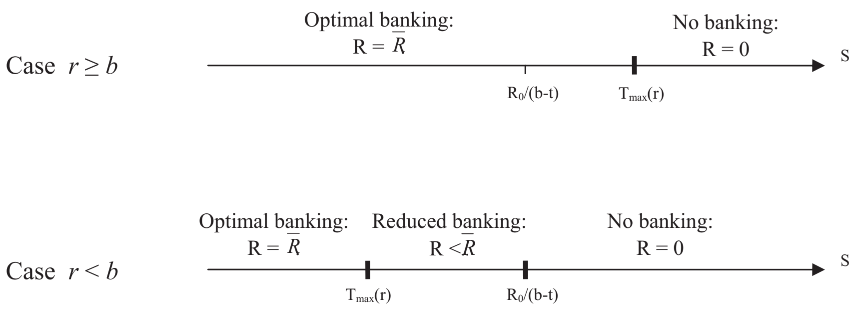

Figure 3. Relationship banking allocation ***R*** as a function of trading opportunities ***S***.

We now consider $r \geq b$. In this case, cutting down on banking when $S > T_{\max}$ leads to a loss in profits that exceeds the capital freed up. This makes the leverage constraint (10) more instead of less binding, reducing the capital available to trading. No smooth reduction in the banking activity can help; in equilibrium, relationship banking unravels. We now have:

$$R = \begin{cases} \overline{R}, & \text{for } S \leq T_{\max} \\ 0, & \text{for } S > T_{\max} \end{cases}, \quad T = \begin{cases} S, & \text{for } S \leq T_{\max} \\ \dfrac{R_0}{b-t}, & \text{for } S > T_{\max} \end{cases}. \tag{12}$$

Overall, higher relationship banking rents r make it less likely that the time-inconsistency problem is binding (i.e., $T_{\max}$ is high). But the consequences of time inconsistency once it is triggered are severe: the credit line business will unravel.

4.2 The Costs of Combining Banking and Trading

Figure 3 shows the volume of the relationship banking activity R under time inconsistency as a function of trading opportunities S. It can help interpret some fundamental changes in the financial sector. Historically, with low scalability of trading (S low) and relatively profitable relationship banking (r high), time inconsistency was not binding because the implicit capital of a relationship bank could accommodate all trading. More recently, the scalability of trading S may have increased and the profitability r of relationship banking declined. Banking is then unable to support the increased volume of trading without triggering detrimental time inconsistency. In Figure 3, this means two things. As long as $r \geq b$ continues to hold (the first case), the no-banking region widens because $T_{\max}$ drops when r falls. Hence, time inconsistency kicks in earlier, and the region where banking continues without distortions due to time inconsistency (the "optimal banking" area) narrows. Together with the reinforcing effect of an increase in S, the likelihood of reaching the "no banking" area grows substantially. Observe that when r drops below b, there is a shift to the lower case in Figure 3. The no banking area widens even more at the expense of the optimal banking area. And even where banking survives it might be distorted (see the "reduced banking" area).[12]

12 The differences between the cases $r \geq b$ and $r < b$ can also be related to industry structure. In a more protected banking system with high r, trading initially poses no problem: banks can safely accommodate a lot of trading ($T_{\max}$ is high). However, once substantial trading opportunities arise,

The underlying trends leading to the deepening of financial markets (higher S) and lower rents in banking (lower r) may well be a reflection of fundamental changes in the economy, particularly developments in information technology come to mind. This also leads to some observations on our distinction between trading and banking. Information technology has enlarged the set of tradeable assets, including now former banking assets. The set of tradeable assets has therefore broadened and the set of banking assets possibly narrowed. Moreover, the degree of specialness of the remaining banking assets might have come down (lower proprietary information content), explaining the lower banking returns r.

In panels A and B of Figure 4 we illustrate for, respectively, the cases $r < b$ and $r \geq b$, the comparative statics of R and T without (as in Section 3) and with time inconsistency.

Proposition 2 (Time inconsistency of capital allocation). *For $\rho < t$, when the profitability of banking is low (r is low) and trading is sufficiently scalable, so that $S > T_{\max}$, the bank allocates insufficient capital to serving the future funding needs of its customers: $R < \overline{R}$. Anticipating this, customers pay lower credit line fees ex ante, and the bank's relationship franchise suffers.*

Proposition 2 is a key result of our model. It provides an important lesson for the industry structure of banking. With limited trading opportunities and relatively profitable banking activities, there are synergies to combining relationship banking and trading. However, more scalable trading coupled with less profitable banking can undermine commitments that are essential for relationship banking. Combining the activities may become costly.

We can now derive bank profits. Recall that the cumulative profit of banking and trading as standalone activities is:

$$\Pi_S = R_0 + r\overline{R}, \tag{13}$$

where $R_0 + r\overline{R}$ is the profit of a standalone relationship bank (see Equation (3)), and the profit of standalone trading (which is not viable) is zero.

The profit of a bank that engages in trading depends on the scalability of trading, S. For $S \leq T_{\max}$, time inconsistency is not present, and the profit is increasing in S:

$$\Pi_C = R_0 + r\overline{R} + tS. \tag{14}$$

For $S > T_{\max}$, time inconsistency distorts the capital allocation, and profit is decreasing in S. Following analysis in Section 4.1, for $r < b$, a bank substitutes trading for banking smoothly, and its profit decreases in a continuous way (use Equation (11)):

$$\Pi_C = \begin{cases} R_0 + \dfrac{rR_0 - Sb(r-t)}{b-r}, & \text{for } T_{\max} < S \leq R_0/(b-t) \\[2ex] R_0 + t\dfrac{R_0}{(b-t)}, & \text{for } S > R_0/(b-t) \end{cases}. \tag{15}$$

the relationship banking arrangements might collapse rapidly. In a more competitive banking system where $r < b$, less trading can be accommodated, but a more smooth transition takes place once trading opportunities increase.

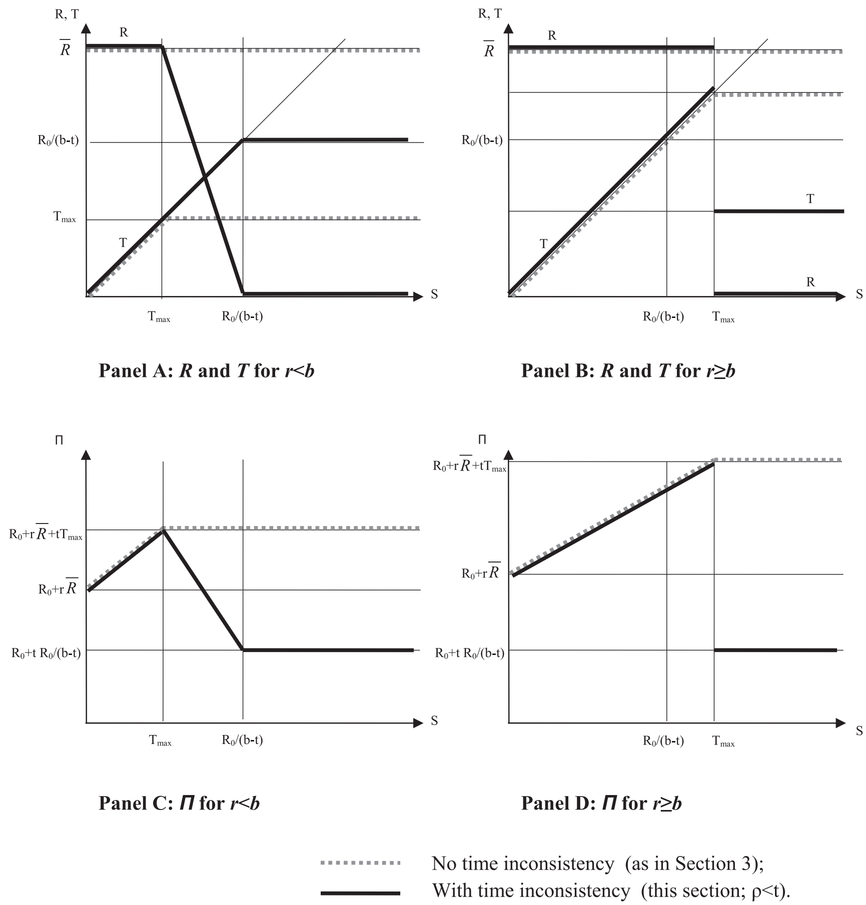

Figure 4. The volumes of relationship banking and trading activities (***R*** and ***T***) and bank profit Π.

For $r \geq b$, the banking activity collapses for $S > T_{\max}$, leading to an immediate drop in profit:

$$\Pi_C = R_0 + t\frac{R_0}{(b-t)}, \text{for } S > T_{\max}. \tag{16}$$

In panels C and D of Figure 4 we illustrate bank profits for the cases $r < b$ and $r \geq b$. Overall, from Equations (14) to (16), in comparison to standalone banking, trading increases bank profit initially (for $S \leq T_{\max}$) but can lead to a loss of profit for $S > T_{\max}$. When a bank fully substitutes trading for relationship banking, so that $R = 0$ and $T = R_0/(b-t)$, its profit Π_C is less than the standalone profit Π_S (compare Equations (13) and (16)) when the value of the relationship banking business linked to funding commitments, $r\overline{R}$, is sufficiently high:

$$r\overline{R} > t\frac{R_0}{(b-t)}. \tag{17}$$

The results can be summarized as follows:

Proposition 3 (Profits under time inconsistency). *For $\rho < t$, the effect of conglomeration on bank profit is inverse U-shaped in trading opportunities S. For low S, $S \leq T_{\max}$, time inconsistency is not present, and profit increases as a bank trades more. For higher S, $S > T_{\max}$, profit falls with additional trading as the time-inconsistency problem intensifies. There exist parameter values such that beyond a certain scale of trading, banks that do not trade generate higher profits than banks that trade.*

Proposition 3 confirms that trading at a large scale can be detrimental for banks.

4.3 Risk Shifting

In the model so far, relationship banking and trading were risk-free. We now extend the model to explore how risk in the trading activity can reinforce the time-inconsistency effects.

Shareholders of a leveraged firm may have incentives for risk shifting when risk is not priced at the margin in the firm's cost of funding (Jensen and Meckling, 1978). This is a standard corporate finance inefficiency that arises when funds are attracted before an investment decision is made to which shareholders cannot commit. Such a lack of commitment, and hence incentives for risk shifting, can be expected in the context of banks too.[13]

Yet the scope to generate risk in a traditional relationship bank might be limited. For example, a bank that has a portfolio of loans with independently distributed returns may have minimal idiosyncratic risk thanks to the law of large numbers, while the aggregate risk (such as interest rate risk) might be absorbed by the bank's charter value if it is sufficiently high thanks to relationship rents. In contrast, banks might be able to use trading to generate highly skewed returns (e.g., from large undiversified positions) and perform risk shifting through trading.

In this section, we examine the setup where relationship banking is safe, while trading can generate skewed returns. Although this pattern need not hold always, it corresponds closely to the experience from the run-up to the recent crisis, when banks accumulated large portfolios of senior tranches of asset-backed securities (mostly real estate loans related) that seemed safe and generated extra return "alpha" in most states of the world, but occasionally led to significant and correlated losses (Acharya *et al*., 2010). Another trading-like way in which banks generated skewed risk was by warehousing transactional loans in order to securitize or syndicate them later, thus being exposed to the inability to do so once syndication and wholesale funding markets "froze" during the crisis (Shin, 2009; Ivashina and Scharfstein, 2010). For this reason, it is instructive to analyze its consequences of such a setup for the interaction between relationship banking and trading activities within a single firm.

Assume that a bank can choose between the safe trading strategy considered before and a risky trading strategy, which for T units invested generates a gross return $(1 + t + \alpha)T$

13 The pricing of risk might be more distorted in banks than in non-financial firms, due to the safety net (deposit insurance and "too-big-to-fail" guarantees). Also, it might be easier for banks than for industrial firms to opportunistically change their risk profile thanks to a higher liquidity of their assets (see Myers and Rajan, 1998). We do not need these additional effects in our model, but if present they would strengthen our results.

with probability p, and 0 with probability $1-p$ (up to the maximum scale of trading S). The binary return of risky trading is a simplification, representing highly skewed outcomes.

We assume that risky trading has a per unit NPV that is lower than that of safe trading, yet positive:

$$0 < p(1+t+\alpha) - 1 < t. \tag{18}$$

Holding the cost of debt fixed, risky trading offers a higher return to shareholders than safe trading (creating incentives for risk shifting), yet not as high as to make the leverage constraint (1) not binding:

$$t < p(t+\alpha) < b. \tag{19}$$

The choice of the trading strategy by the bank is not verifiable.

In the model so far, the interest rate on bank borrowing was zero because the bank was risk-free. Yet, with risky trading, a bank may be unable to repay its creditors in full. Accordingly, creditors *ex ante* set the interest rate i based on their expectations of the bank's future trading strategy. For simplicity, we assume that if multiple equilibria are possible, the creditors set the lower rate. The bank chooses safe trading when indifferent. We focus on the richest case that combines time inconsistency ($\rho < t$) and a smooth contraction of banking ($b > r$).

We can prove the following result:

Proposition 4 (Risk shifting). *When the profitability of banking is low (r is low) and trading is sufficiently scalable (S is high), the bank engages in risky trading as a form of risk shifting. The bank internalizes the costs of risk shifting through higher borrowing costs and lower borrowing capacity. Risk shifting reduces bank profits, makes the bank more credit-constrained, and in this way damages the bank's relationship franchise (its ability to serve relationship customers).*

Proof. See Appendix A.

The proposition implies that a bank only uses trading for risk shifting when the profitability of relationship banking is relatively low, while the scalability of trading is high. The intuition is that, from the perspective of bank shareholders, risky trading has a fixed cost—the loss of the relationship banking franchise value in bankruptcy with probability $(1-p)$. The benefits of risky trading—the additional return α—are proportional to the scale of trading. Hence, the scale of risky trading has to be high enough to compensate for putting the relationship banking franchise at risk.

Note that the shift to risky trading has two negative effects. First, it has a direct negative impact on bank profits because risky trading has a lower NPV than safe trading. Although shareholders do not internalize this *ex post*, they internalize it *ex ante* through higher interest rates on bank borrowing. Second, lower bank profits reduce bank borrowing capacity, and this further undermines bank profitability.

It is also useful to note that the two inefficiencies—time inconsistency of capital allocation and risk shifting—can reinforce each other, so that the presence of one makes the presence of the other more likely.

Effect 1. *Risk shifting makes time inconsistency in capital allocation more likely by increasing the ex post return on risky trading. There exist parameter values:*

$t < \rho < p(t + \alpha)$, such that there is no time inconsistency in the absence of risk shifting ($t < \rho$), but it is present under risk shifting ($\rho < p(t + \alpha)$).

Effect 2. *Time inconsistency induces risk shifting by increasing the equilibrium scale of trading*. This follows from Propositions 2 and 4. *Risky trading is only optimal when the scale of trading is sufficiently high, and time inconsistency increases the equilibrium volume of trading.*

Effect 3. *Time inconsistency induces risk shifting by reducing the relationship bank's franchise value. The incentives for risk shifting are countered by the risk of a loss of the bank's relationship franchise value with probability $(1 - p)$. Time inconsistency reduces that franchise value, and hence lowers the cost of risk shifting. We offer in Appendix B an extension of the model that captures this effect.*

5. Discussion

This section discusses modeling features, and summarizes empirical and policy implications.

5.1 Front-loaded Income in Relationship Banking

We first highlight the relevance of our approach to modeling relationship banking as a funding insurance activity. There is ample evidence that banks play a substantial role in providing funding insurance to customers. For example, credit lines represent up to 70% of bank lending (Kashyap, Rajan, and Stein, 2002; Berger and Bouwman, 2009). Importantly, our credit line setup can be seen also as a formalization of a broader variety of circumstances where relationship banks offer *implicit* funding commitments. A notable example is a bank's presence in local markets. The literature shows that a credible local market presence goes hand in hand with an implicit commitment to facilitate the funding of customers, especially in times of economic stress (Petersen and Rajan, 1994; Bae, Kang, and Lim, 2002; Beck *et al.*, 2013). In return, local banks can charge customers higher fees in normal times (Bolton *et al.*, 2013; Sette and Gobbi, 2015). Such an arrangement is only possible when a bank's continued local market presence is credible. Our analysis highlights that time-inconsistency issues involving excessive trading can undermine this.

Our emphasis on funding insurance as a critical feature of relationship banking contrasts with the commonly postulated approach of describing long-term relationships between a bank and its customers based on information capture. There, borrowers are subsidized initially, while hold-up allows the bank to recoup the subsidies later (Petersen and Rajan, 1995). This leads to back-loaded income, making relationship banking highly attractive *ex post*. Observe that the *ex post* rents in banking have declined in the recent past as higher competition and more easily available borrower information have eroded the informational advantage of banks. Accordingly, the informational capture aspect of relationship banking may have become less important.

5.2 Commitment Issues

Our analysis makes an important assumption that banks cannot commit to limit the scale of trading. A rationale is that the verification of such a private commitment might be difficult. Banks may always have to engage in some trading to support their lending activity (e.g., to hedge lending exposures), so committing not to trade at all might be prohibitively costly. And once banks trade, trading may take many forms, especially during times of rapid financial innovation. This makes it hard to write *ex ante* contracts on the volume or nature of trading.[14]

Another issue is whether the time-inconsistency problem in internal capital allocation could be mitigated by returning the capital back to shareholders via dividends or share buy-backs. Although this looks promising, returning capital to shareholders would also undermine the relationship banking franchise. To understand this, note that in the normal course of business a relationship bank should always maintain "unused" capital (spare borrowing capacity) in order to cover future funding needs of customers. The model highlights that such capital can be misallocated to trading, making the bank unable to fulfill its relationship commitments. If this unused capital was returned to shareholders, the bank would still be unable to make good on its lending commitments. So the relationship franchise would suffer in either case.

The problem of time inconsistency between short- and long-term activities that is the focus of this article is reminiscent of the literature that shows how trading at an intermediate stage in dynamic models of financial intermediation might undermine commitment. For example, Jacklin (1987) shows that trading possibilities may undermine the intertemporal smoothing between early and late consumers in the Diamond and Dybvig (1983) analysis, while Bhide (1993) points at a lack of shareholder discipline under diffused (and liquid) ownership.

5.3 The Discretionary Credit Line Contract

This section presents a stylized model to microfound the discretionary credit line contract in Section 4. We establish why a credit line may add value and explain the discretion feature.

Consider the following framework. At date 0, the firm is endowed with a project that can produce X at date 2. At date 1, there are two states of the world, H and L (with probabilities $p_h = 1 - p_l$ and p_l, respectively). In state L, the firm needs to invest an additional 1 unit for the project to succeed and realize X for sure (otherwise it realizes zero); in state H no such investment is required and X will be realized. The investment that might be needed at date 1 is externally funded. The amount $\Omega < X$ is pledgeable to financiers. We let $1 < \Omega < X$, so that project continuation is optimal.

Assume that the firm has a relationship with one bank and cannot borrow elsewhere. Hence, the bank has all the bargaining power vis-à-vis the firm. We introduce the following friction. The amount $\Omega < X$ is pledgeable only if the bank is engaged with the firm starting from date 0. (E.g., the bank needs to continuously maintain and monitor the firm's transaction accounts to collect any repayments.) This friction implies that at date 0 the bank can make the firm a "take-it-or-leave-it" offer, effectively forcing the firm to buy a credit line for the potential funding need of 1 at date 1. The benefit of the credit line to the bank is that it can extract higher rents from the firm. To see this, note that without a credit line the bank can only charge the firm Ω upon lending 1 in state L, leaving the firm with rents $X - \Omega$. With a credit line, the bank contracts at date 0 and can ask for extra compensation such that in the expected value sense the firm is giving up all rents from state L. The bank

14 One could envision a bimodal industry structure where some banks exploit trading opportunities while others abstain from trading altogether. But, as indicated above, such total avoidance of trading has costs—for example, inability to hedge lending exposures, and therefore may be detrimental to relationship banking as well.

cannot ask for more than that, because otherwise the firm would be better off without a credit line, abandoning the project in state L.

Under a credit line, the bank charges the firm an unconditional fee F payable in both states,[15] and a repayment obligation $\Omega - F$ upon the borrowing of 1 at date 1. The firm's participation constraint is given by:

$$(1 - p_l)X + p_l(X - (\Omega - F)) - F = (1 - p_l)X, \tag{20}$$

where $(1 - p_l)X$ is the firm's profit in state H, $p_l(X - (\Omega - F))$ is the expected repayment on borrowing in state L, and F is the credit line fee. Thus,

$$F = (X - \Omega)\frac{p_L}{1 - p_L}.$$

In total, the firm pays the bank F in state H and Ω in state L.

In principle, there exist other contracts—with a higher credit line fee F and a lower repayment on borrowing $\Omega - F$ (i.e., with repayments more loaded on state H)—that generate the same expected payoffs for the firm and the bank as the contract described above. But when the bank has discretion on whether to make good on lending commitments (a feature that we will explain momentarily), the payments schedule with the highest possible repayment in state L is the only one that is renegotiation-proof.[16]

We can now motivate the need for the bank's discretion. Assume that after the credit line is contracted, but before borrowing, the firm can choose (at no cost) to transform its assets such that they are no longer pledgeable to the bank.[17] Then, the bank will never offer a non-discretionary credit line; absent discretion the firm would always transform its assets. Allowing the bank discretion makes the credit line feasible.

As a final step, consider the effect of trading opportunities for the bank. Assume that, at date 1 in state L, the bank with probability π gets an opportunity to invest 1 unit in trading to produce t, where $\Omega < t < X$ (with probability $1 - \pi$ no trading opportunity exists). When the bank invests in trading, it cannot simultaneously lend to the firm.

15 In a richer model where a firm has some initial monetary endowment we could have had a fee being paid at date 0.

16 To see this, note that the bank's discretion on whether to lend to the firm after the credit line has been agreed creates a point of renegotiation between the bank and the firm. Assume that the firm has purchased a credit line from the bank at date 0. Consider the case where, at date 1, the state L is realized (and the firm faces a funding need). At that point, the bank can threaten to withhold funding from the firm unless the firm agrees to increase the repayment to the bank. Specifically, assume that, at that point, the bank can make a take-it-or-leave-it offer to the firm (i.e., as before, the bank has all the bargaining power). Whenever the total payment to the bank according to the credit line contract in state L is less than Ω, the bank can demand from the firm the highest possible payment, Ω, or else refuse to lend. If the firm agrees, it will keep the non-pledgeable part of profits, $X - \Omega$. If the firm declines, it will have to liquidate the project and obtain zero. Accordingly, the firm will always agree to the bank's offer. Thus, only the contract with a maximum possible payment Ω to the bank in state L is renegotiation-proof.

17 This possible transformation captures in a simplified way a variety of borrower moral hazard behaviors that may arise under fully committed funding insurance. For example Acharya et al. (2014a) suggest that once full insurance is in place, firms could engage in risky investments that increase the likelihood of liquidity shocks that they are insured against. The bottom line is that borrower moral hazard could undermine the feasibility of funding insurance.

If the bank could commit to lend to the firm through a non-discretionary credit line (and abstain from trading), it would have done so because lending is more profitable than trading ($t < X$). But the credit line can only be discretionary (see above). Discretion makes the bank choose *ex post* to trade whenever such an opportunity is present because $\Omega < t$. The bank then can only lend to the firm with probability $p_L(1-\pi)$. This reduces the equilibrium credit line fee to:

$$F' = (X - \Omega)\frac{p_L(1-\pi)}{1 - p_L(1-\pi)} < F, \tag{21}$$

and expected bank profit to: $\{(X-1)(1-\pi) + (t-1)\}\pi p_L$. This is strictly less than the profit in the absence of trading opportunities, $(X-1)p_L$, because $t < X$.

5.4 Credit Constraints in Trading

Our analysis of credit constraints for banks builds on Holmstrom and Tirole (1998) formulation where a firm's borrowing capacity is limited to a multiple of the owner-manager's net worth. As discussed in Section 2, this formulation can be rationalized by allowing managers to convert assets into private benefits through absconding or shirking (i.e., consuming perquisites instead of working). Either action creates private benefits proportional to asset size, but makes the assets themselves worthless for the firm (see Equation (1)). We have further assumed that the intensity of the credit rationing problem is identical for banking and trading activities: the parameter b is the same in Equation (2) and in Equation (5).

In practice, the leverage constraint in trading might be different from that in banking. The higher liquidity of trading assets might induce transformation risk (Myers and Rajan, 1998) and tighten the credit constraint relative to banking. A counter effect is that one can possibly easier pledge trading assets as collateral. The latter would loosen the leverage constraint. A looser credit constraint in trading could impact our results. To see this, assume that the absconding problem in trading is softer than that in banking. That implies that in Equation (5) the bank faces a lower value of b, call this b', with $b' < b$. If pledging is very effective, so that $b' < t$, standalone trading is possible. However, pledging may only be partially effective in loosening the leverage constraint. For example, pledging can be costly (think of custody costs or the higher costs of adjusting the portfolio when assets are pledged), leading to lower returns on trading t that offset the beneficial effects of a lower b in Equation (5). When pledging is only partially effective, so that $t < b' < b$, the leverage constraint is looser than that in banking, but standalone trading is still not possible. Allowing for such b' transforms the leverage constraint for the bank (7) to:

$$R_0 + rR + tT \geq bR + b'T, \tag{22}$$

and increases the maximum scale of trading that does not trigger time inconsistency (see Equation (8)) to:

$$T'_{\max} = \frac{R_0 + \overline{R}(r-b)}{b'-t} > T_{\max}. \tag{23}$$

Here, for sufficiently high S, such that $S > T'_{\max}$, our results on time-inconsistency problems in combining banking and trading persist.

5.5 Empirical Implications

The analysis leads to several implications that are useful for understanding the dynamics of banks that engage in trading. The most direct implications are as follows:

Trading by banks and bank profitability. While trading at a limited scale could enhance bank profitability, sizable trading can reduce bank profits. That is, banks that trade at a large scale undermine their relationship banking franchise. As a consequence, profits from repeated or nearby customers (both proxies for the presence of relationships) may decline, and the value of the intangibles associated with such customers may suffer. Relationship-oriented banks may suffer significant diseconomies of scope.

Impact on firms. When banks trade more, the benefits of bank relationships for firms decline. Firms would derive less value from proximity to a bank, or from having a credit line. To compensate for a higher probability of a bank reneging on funding commitments, firms may use more cash in their liquidity management. To the extent that banks cushion firm funding costs during downturns, trading by banks may contribute to macroeconomic cyclicality.

Impact on contractual arrangements. The possibility of trading may force banks to make their funding commitments less discretionary, that is, with less discretion in credit lines and more formal arrangements with local market customers. But to the extent that discretion in funding insurance is desirable (e.g., to counter borrower moral hazard), the volume of funding insurance provided by banks may decline.

There are also some broader implications:

Why is trading by banks a bigger issue today that it was historically? Financial development has undermined banking through two channels: deepening of financial markets (with more scalable trading) and less profitable relationship banking (due to higher competition and more widely available customer information). Both increase bank incentives to over-allocate capital to trading.

Challenges for universal banks. A broad implication from the increased severity of the time-inconsistency problem is that combining banking and trading (the traditional European universal bank model shared by some US conglomerates) might have become less sustainable. Universal banks have historically combined a sizable relationship banking activity with a much smaller transaction-based activity. Now banks might allocate too many resources to trading, leading to lower profits and higher risk.

Role of credit ratings. Anticipating a bank's ability to honor commitments has become more important. Credit ratings may in part reflect the risk bearing capacity that will not be filled up opportunistically. Hence, borrowers may anticipate that banks benefitting from a higher rating are better able to deliver on (implicit) guarantees of future funding. This could help explain the importance of ratings in the financial services industry.

Too-big-to-fail and procyclicality. The deepening of financial markets allows banks to engage in trading on a larger scale, increasing bank size and complexity (amplifying too-big-to-fail problems). Also, banks' exposure to boom-and-bust patterns in financial markets might be reinforced by the cyclicality of time-inconsistency problems (i.e., elevated trading activity during financial market upswings). Therefore, trading by banks, procyclicality and "too-big-to-fail" are interrelated.

5.6 Policy Implications

Our analysis is partial equilibrium, so we need to be cautious about drawing strong implications. Still, we have highlighted specific distortions, that is, over-allocation of capital to trading and the use of trading for risk shifting. Accordingly, we can consider how some regulatory proposals that deal with trading by banks may help correct these distortions.

Capital charges on banks' trading assets. Increased capital charges on trading assets can discourage excess trading by reducing the return to trading. Once the return to trading t falls below the *ex post* return to banking ρ, the time-inconsistency problem in capital allocation disappears.

Restricting trading. We indicated in Section 5.2 that a private commitment to limit trading might be impossible. Public regulation might be difficult as well, but if feasible could increase bank value (Proposition 3) and reduce bank risk (Proposition 4).

Segregation. Segregating trading into a separate subsidiary could prevent risk bearing capacity in the relationship bank from being used to underwrite trading. However, it is crucial not only to block recourse but also to limit capital transfers to the trading subsidiary to minimize the time-inconsistency problem in capital allocation.

Our analysis also offers insights into some additional policy questions:

Can trading move to "the shadow"? A common argument against stronger bank regulation is that trading may move to the unregulated sector and become an even greater concern. Our analysis suggests that such a migration of trading activities would not necessarily occur. It is a bank's franchise value that induces trading at possibly too large a scale. Standalone trading would be capital constrained and have lower scale, hence pose lower risks to financial stability.

Investment banking: standalone or within bank groups? The crisis has highlighted the instability of standalone investment banks. Our analysis suggests that this may be because investment banks expanded their trading operations to a scale that was not compatible with the capital constraints that trading should face. Specifically, they may have substituted relationship-oriented activities that generate implicit capital (e.g., underwriting) for trading.[18] Investment banks may have been able to operate at such scale thanks to perceived government guarantees, or because the market overlooked their transition from a relationship-oriented to a trading-dominated business.

Government policy vis-a-vis standalone investment banks. When markets started doubting the viability of standalone investment banks, the US authorities adopted two approaches to rescuing them. One approach was to merge investment banks with commercial banks (as in the case of Bear Stearns and Merrill Lynch). The other approach was to grant them commercial bank licenses and hence access to discount window and other central bank facilities (e.g., Goldman Sachs and Morgan Stanley). Our analysis suggests that while a merger with a commercial bank might be a quick way to inject implicit capital and stabilize the trading operation, in the medium term such conglomeration may induce time-inconsistency problems.[19]

18 Some investment banks also have lost risk bearing capacity intrinsic to partnerships by abandoning their partnership structure (Morrison and Wilhelm, 2008).

19 An example illustrating this point is that following the creation of Bank of America Merrill Lynch, analysts and regulators expressed worry that the trading exposures of Merrill Lynch may become "a drain on the resources" of Bank of America. See "BofA Said to Split Regulators Over Moving Merrill Derivatives to Bank Unit," Bloomberg, December 18, 2011.

Dynamic approach to bank regulation. A regulatory framework that focuses only on contemporaneous bank risk may insufficiently capture bank incentives to take risks in the future. Regulation should be cognizant of time-inconsistency problems. The proliferation of financial markets has made these concerns more acute.

6. Conclusion

The article studies the interaction between long-term relationship banking and short-term trading-like activities of banks. The increased involvement of banks in trading is a fundamental change affecting the industry over the last decades. However, banks may over-allocate capital to trading due to time-inconsistency problems, and in doing so undermine their relationship banking franchise. This effect is reinforced when banks can use trading for risk shifting. The distortions are more likely when the scalability of trading is high and the profitability of relationship banking is low. These factors were in play in recent decades, as information technology led to deeper financial markets and weakened the grip of banks over their relationship borrowers. As a result, trading by banks, while benign historically, might have recently become distortive.

On a fundamental level, the paper asks questions about the essence of the modern banks' business model. One could argue that recent financial development has led to an "identity crisis" in banking. Many of the activities that create franchise value for banks involve long-term commitments to customers. Particularly, the provision of funding insurance, via credit lines or through implicit commitments to local markets, is core to the business of banking. But access to transactional, opportunistic trading activities may undermine such commitments. Our partial equilibrium analysis cannot give absolute answers, but it points to a range of distortions that bank managers and regulators could address to strengthen the business models of banks.

Appendix A: Proof of Proposition 4

We first derive conditions under which the bank shifts to risky trading. To do so, start by assuming that no default risk is priced in. This allows us to derive the bank's risk choice.

Consider the payoff to risky trading under $i = 0$. When a bank invests R in relationship banking and T in risky trading, it obtains at date 2 from relationship banking $R_0 + (1 + r)R$ and from trading $(1 + t + \alpha)T$ with probability p and 0 otherwise. The bank has to repay creditors $R + T$. It can do so in full when trading succeeds. When trading produces a zero return, the bank has sufficient returns to repay in full only if $R_0 + (1 + r)R \geq R + T$, corresponding to an upper bound on the volume of trading:

$$T \leq R_0 + rR. \tag{A.1}$$

When Equation (A.1) holds, bank debt is safe, and shareholders fully internalize the losses from risky trading. Their payoff:

$$\Pi_{\text{Risky}}|_{T \leq R_0 + rR} = R_0 + rR + p(1 + t + \alpha)T - T \tag{A.2}$$

is lower than the payoff with safe trading: $\Pi_{\text{Risky}}|_{T \leq R_0 + rR} < \Pi_C$, where Π_C is given in Equation (6), because risky trading has a lower NPV than safe trading (see Equation (18)). Hence, bank shareholders never choose risky trading if they fully internalize possible losses.

We thus focus on the case $T > R_0 + rR$. Here, there is default risk: a bank cannot repay creditors in full when the risky trading strategy produces a zero return. The shareholders' return is:

$$\Pi_{\text{Risky}} = p(R_0 + rR + (t+\alpha)T). \tag{A.3}$$

The bank chooses risky trading when $\Pi_{\text{Risky}} > \Pi_C$. From Equations (6) and (A.3), this holds when the scale of trading exceeds a threshold T_{Risky} such that:

$$T > T_{\text{Risky}} = \frac{(R_0 + rR)(1-p)}{p(t+\alpha) - t}. \tag{A.4}$$

To express T_{Risky} in exogenous variables, note that when a bank allocates T to trading, the borrowing capacity left to banking (from Equation (10)) is:

$$R = \frac{R_0 - T(b-t)}{b-r}. \tag{A.5}$$

Substituting this into Equation (A.4) gives:

$$T_{\text{Risky}} = \frac{R_0(1-p)}{-r(p\alpha - b(1-p))/b + (p\alpha - t(1-p))}. \tag{A.6}$$

T_{Risky} is a key threshold of this model; it shows the minimum scale of trading that induces risk shifting. We can show the following:

Observation 1. $\partial T_{\text{Risky}}/\partial\alpha < 0$ and $\partial T_{\text{Risky}}/\partial p < 0$: when risky trading has a higher upside α and/or a higher probability of success p, the switch to risky trading occurs at a lower scale of trading.

Proof: By differentiation:

$$\partial T_{\text{Risky}}/\partial\alpha = bp(r-b)\frac{R_0(1-p)}{(-r(p\alpha - b(1-p)) + b(p\alpha - t(1-p)))^2} < 0,$$

and

$$\partial T_{\text{Risky}}/\partial p = b\alpha R_0 \frac{r-b}{(-r(p\alpha - b(1-p)) + b(p\alpha - t(1-p)))^2} < 0,$$

because $r < b$.

Observation 2. $\partial T_{\text{Risky}}/\partial R_0 > 0$ and $\partial T_{\text{Risky}}/\partial r > 0$: when the value of the relationship banking franchise is higher, the switch to risky trading occurs at a higher scale of trading.

Proof: Also by differentiation:

$$\partial T_{\text{Risky}}/\partial R_0 = \frac{(1-p)}{-r(p\alpha - b(1-p))/b + (p\alpha - t(1-p))} > 0.$$

This is positive because:

$$-r(p\alpha - b(1-p))/b + (p\alpha - t(1-p)) = [(b-r)p\alpha + (1-p)b(r-t)]/b > 0$$

and

$$\partial T_{\text{Risky}}/\partial r = bR_0(1-p)\frac{p\alpha - b(1-p)}{\left(-r(p\alpha - b(1-p)) + b(p\alpha - t(1-p))\right)^2} > 0$$

which holds because

$$p\alpha - b(1-p) > p\alpha - p(t+\alpha)(1-p) = p(p(t+\alpha)-t) > 0.$$

We can now derive the interest rates, capital allocation, and profits under risky trading. To limit attention to the more insightful case, we let:

$$T_{\max} < T_{\text{Risky}} < \frac{R_0}{b-(p(1+t+\alpha)-1)}. \tag{A.7}$$

This ensures that (*i*) risk shifting only occurs when the time-inconsistency problem in capital allocation is already present and (*ii*) implicit capital is sufficient to support the scale of trading necessary for risk shifting.

For $S \le T_{\text{Risky}}$, the bank chooses the safe trading strategy with the allocation of borrowing capacity given by Equation (11); for $S > T_{\text{Risky}}$, the bank chooses risky trading. Note that when there is time inconsistency in the capital allocation under safe trading ($\rho < t$), it also exists under risky trading because the *ex post* return to shareholders under risky trading is higher than that under safe trading, that is, $\rho < t < p(t+\alpha)$ (see Equation (19)). Therefore, under risky trading, the bank will also first allocate the borrowing capacity to trading before using the remainder for banking. The interest rate i follows from the creditors' zero-profit condition:

$$(R+T) = p(R+T)(1+i) + (1-p)(R_0 + (1+r)R), \tag{A.8}$$

where $(R+T)$ is the amount borrowed by the bank, $(R+T)(1+i)$ is the debt repayment when risky trading succeeds with probability p, and $(R_0 + (1+r)R)$ is the value of bank assets transferred to creditors in bankruptcy with probability $(1-p)$. The capacity R left for the banking activity after allocating T to trading follows from the IC condition (leverage constraint):

$$p(R_0 + (r-i)R + (t+\alpha-i)T) \ge b(T+R), \tag{A.9}$$

where the left-hand side is the expected payoff to bank shareholders in normal operations, and the right-hand side is the moral hazard payoff.

Similar to Equation (11), use Equations (A.8) and (A.9) as equality to obtain the allocation of borrowing capacity:

$$R = \begin{cases} \dfrac{R_0 - S((1-p)+b-p(t+\alpha))}{b-r}, \ \text{for } T_{\text{Risky}} < S \le \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \\ 0, \ \text{for } S > \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \end{cases}, \tag{A.10}$$

$$T = \begin{cases} S, \ \text{for } T_{\text{Risky}} < S \le \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \\ \dfrac{R_0}{b-(p(1+t+\alpha)-1)}, \ \text{for } S > \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \end{cases}.$$

Similarly to Equation (15), bank profits are:

$$\Pi_{\text{Risky}} = \begin{cases} p\left(R_0 + \dfrac{rR_0 - Sb(r-(p(1+t+\alpha)-1))}{b-r}\right), & \text{for } T_{\text{Risky}} < S \leq \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \\ p\left(R_0 + \dfrac{(p(1+t+\alpha)-1)R_0}{b-(p(1+t+\alpha)-1)}\right), & \text{for } S > \dfrac{R_0}{b-(p(1+t+\alpha)-1)} \end{cases}. \tag{A.11}$$

Appendix B: Extension—Time Inconsistency, Franchise Value, and Risk Shifting

We offer an extension capturing the fact that the presence of time inconsistency may induce risk shifting because it lowers the bank's franchise value (Effect 3 in Section 4.3). To show this interaction, we need to enrich the model in order to isolate the effect of time inconsistency on franchise value. Assume that the presence of the time-inconsistency problem is uncertain at date 0. At date 0 let there be a probability γ that at date 1, $t = t_{\text{Low}} < \rho$, so that there is no time inconsistency; and a probability $1-\gamma$ that $t = t_{\text{High}} > \rho$, and hence time inconsistency is present. A different probability $(1-\gamma)$ of the presence of time inconsistency affects a bank's franchise value, and through it the threshold value T_{Risky} (see Equation (A.6)).

Consider the realization $t = t_{\text{Low}}$ at date 1, so that time inconsistency is present. Then, at date 1, the payoff to the bank's shareholders from safe trading is:

$$\Pi_C^{\gamma} = R_0 + (r-\rho)\left(\gamma\overline{R} + (1-\gamma)R_{\text{anticipated}}\right) + \rho R + tT. \tag{A.12}$$

The expression Π_C^{γ} is similar to Π_C in Equation (9), except for the second term. Time inconsistency now arises with probability $1-\gamma$ rather than with probability 1, hence compared with Equation (9) the credit line fees (and the bank's franchise value) are higher by $(r-\rho)\gamma\left(\overline{R} - R_{\text{anticipated}}\right)$.

Similarly, the payoff from risky trading is:

$$\Pi_R^{\gamma} = p\left(R_0 + (r-\rho)\left(\gamma\overline{R} + (1-\gamma)R_{\text{anticipated}}\right) + \rho R + (t+\alpha)T\right). \tag{A.13}$$

Setting $R = R_{\text{anticipated}}$ and equating Π_C^{γ} and Π_R^{γ} provides the threshold for a switch to risk shifting (similar to Equation (A.4)):

$$T_{\text{Risky}}^{\gamma} = \frac{\left(R_0 + (r-\rho)\gamma\left(\overline{R} - R\right) + rR\right)(1-p))}{p(t+\alpha)-t}. \tag{A.14}$$

Observe that a lower γ—meaning more time inconsistency—lowers $T_{\text{Risky}}^{\gamma}$, hence risk shifting becomes more likely when time inconsistency intensifies.

References

Acharya, V. V., Almeida, H., Ippolito, F., and Perez, A. (2014a) Credit lines as monitored liquidity insurance: Theory and evidence, *Journal of Financial Economics* **112**, 287–319.

Acharya, V. V., Almeida, H., Ippolito, F., and Perez, A. (2014b) A new perspective on bank-dependency: The liquidity insurance channel. *ECB working paper 1702*.

Acharya, V. V., Cooley, T. F., Richardson, M. P., and Walter, I. (2010) Manufacturing tail risk: A perspective on the financial crisis of 2007–09, *Foundations and Trends in Finance* **4**.

Acharya, V. V., Pagano, M., and Volpin, P. (2013) Seeking alpha: Excess risk taking and competition for managerial talent. *NBER working paper 18891*.

Adrian, T. and Shin, H. S. (2013) Procyclical leverage and value-at-risk, *Review of Financial Studies* **27**, 373–403.

Bae, H.-K., Kang, J.-K. and Lim, C.-W. (2002) The value of durable bank relationships: Evidence from Korean banking shocks, *Journal of Financial Economics* **64**, 181–214.

Baele, L., De Jonghe, O. and Vander, Vennet, R. (2007) Does the stock market value bank diversification? *Journal of Banking and Finance* **31**, 1999–2023.

Beck, T., Degryse, H., De Haas, R. and van Horen, N. (2013) When arm's length is too far: Relationship banking over the business cycle. *Unpublished working paper*.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2010) Financial institutions and markets across countries and over time: The updated financial development and structure database, *World Bank Economic Review* **24**, 77–92.

Berger, A. N. and Bouwman, C. H. S. (2009) Bank liquidity creation, *Review of Financial Studies* **22**, 3779–3837.

Bhide, A. (1993) The hidden cost of stock market liquidity, *Journal of Financial Economics* **34**, 31–51.

Bolton, P., Freixas, X., Gambacorta, L., and Mistrulli, P. E. (2013) Relationship and transaction lending in a crisis. *NBER working paper 19467*.

Boot, A. W. A., Greenbaum, S. I., and Thakor, A. V. (1993) Reputation and discretion in financial contracting, *American Economic Review* **83**, 1165–1183.

Boot, A. W. A. and Thakor, A. V. (2000) Can relationship banking survive competition? *Journal of Finance* **55**, 679–713.

Boyd, J., Chang, C., and Smith, B. D. (1998) Moral hazard under commercial and universal banking, *Journal of Money, Credit and Banking* **30**, 426–468.

Brunnermeier, M., Dong, G., and Palia, D. (2012) Banks' non-interest income and systemic risk. Unpublished working paper.

Calomiris, C. and Kahn, C. (1991) The role of demandable debt in structuring optimal banking arrangements, *American Economic Review* **81**, 497–513.

Chakraborty, I., Goldstein, I., and MacKinlay, A. (2014) Do asset price booms have negative real effects? Unpublished working paper.

Chava, S. and Roberts, M. (2008) How does financing impact investment? The role of debt covenants, *Journal of Finance* **63**, 2085–2121.

De Jonghe, O. (2010) Back to basics in banking? A micro-analysis of banking system stability, *Journal of Financial Intermediation* **19**, 387–417.

Dell'Ariccia, G. and Marquez, R. (2006) Lending booms and lending standards, *Journal of Finance* **61**, 2511–2546.

Demirguç-Kunt, A. and Huizinga, H. (2010) Bank activity and funding strategies: The impact on risk and returns, *Journal of Financial Economics* **98**, 626–650.

Demiroglu, C. and James, C. (2011) The use of bank lines of credit in corporate liquidity risk management: A review of empirical evidence, *Journal of Banking and Finance* **35**, 775–782.

DeYoung, R. and Rice, T. (2004) Noninterest income and financial performance at US commercial banks, *Financial Review* **39**, 101–127.

DeYoung, R. and Torna, G. (2013) Nontraditional banking activities and bank failures during the financial crisis, *Journal of Financial Intermediation* **22**, 397–421.

Diamond, D. W. and Dybvig, P. H. (1983) Bank runs, deposit insurance and liquidity, *Journal of Political Economy* **91**, 401–419.

Donaldson, G. (1984) *Managing Corporate Wealth*, Praeger Publishers, New York.

Edwards, F. R. and Mishkin, F. S. (1995) The decline of traditional banking: Implications for financial stability and regulatory policy, *FRBNY Economic Policy Review* **July**, 27–47.

Fahlenbrach, R., Prilmeier, R., and Stiltz, R. (2012) This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis, *Journal of Finance* **67**, 2139–2185.

Fang, L., Ivashina, V., and Lerner, J. (2010) "An unfair advantage"? Combining banking with private equity investing. *Unpublished working paper*.

Freixas, X., Lóránth, G., and Morrison, A. D. (2007) Regulating financial conglomerates, *Journal of Financial Intermediation* **16**, 479–514.

Gale, D. (1993) Informational capacity and financial collapse, in C. Mayer and X. Vives (eds), *Financial Intermediation in the Reconstruction of Europe*, Cambridge University Press, Cambridge, UK, pp. 117–147.

Hauswald, R. and Marquez, R. (2006) Competition and strategic information acquisition in credit markets, *Review of Financial Studies* **19**, 967–1000.

Holmstrom, B. and Tirole, J. (1998) Private and public supply of liquidity, *Journal of Political Economy* **106**, 1–40.

Ivashina, V. and Scharfstein, D. (2010) Bank lending during the financial crisis of 2008, *Journal of Financial Economics* **97**, 319–338.

Jacklin, C. (1987) Demand deposits, trading restrictions, and risk sharing, in: E. D. Prescott and N. Wallace (eds), *Contractual Arrangements for Intertemporal Trade*, University of Minneapolis Press, Minneapolis, pp. 26–47.

Jensen, M. C. and Meckling, W. H. (1978) Theory of the firm: Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics* **3**, 305–360.

Kahn, C., and Winton, A. (2004) Moral hazard and optimal subsidiary structure for financial institutions, *Journal of Finance* **59**, 2531–2575.

Kashyap, A. K., Rajan, R., and Stein, J. S. (2002) Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking, *Journal of Finance* **57**, 33–73.

Keeley, M. C. (1990) Deposit insurance, risk, and market power in banking, *American Economic Review* **80**, 1183–1200.

Kroszner R. and Rajan, R. J. (1994) Is the Glass-Steagall Act justified?: Evidence from the U.S. experience with universal banking 1921–1933, *American Economic Review* **84**, 810–832.

Kyle, A. (1985) Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335.

Laeven, L. and Levine, R. (2007) Is there a diversification discount in financial conglomerates? *Journal of Financial Economics* **85**, 331–367.

Liikanen Report. (2012) The high-level expert group on reforming the structure of the EU banking sector: Final report. Available at: http://ec.europa.eu/internal_market/bank/docs/high-level_expert_group/report_en.pdf.

Mercieca, S., Schaeck, K., and Wolfe, S. (2007) Small European banks: Benefits from diversification? *Journal of Banking and Finance* **31**, 1975–1998.

Morrison, A. D. and Wilhelm, W. J. (2008) The demise of investment banking partnerships: Theory and evidence, *Journal of Finance* **63**, 311–350.

Myers, S. C. and Rajan, R. G. (1998) The paradox of liquidity, *Quarterly Journal of Economics* **113**, 733–771.

NY Fed [Federal Reserve Bank of New York]. (2014) Quarterly trends for consolidated U.S. banking organizations. Available at: https://www.newyorkfed.org/research/banking_research/quarterly_trends.html.

Petersen, M. A. and Rajan, R. G. (1994) The benefits of lending relationships: Evidence from small business data, *Journal of Finance* **49**, 3–37.

Petersen, M. A. and Rajan, R. G. (1995) The effect of credit market competition on lending relationships, *Quarterly Journal of Economics* **110**, 407–443.

Puri, M. (1996) Commercial banks in investment banking: Conflict of interest or certification role? *Journal of Financial Economics* **40**, 373–401.

Rajan, R. J., Servaes, H., and Zingales, L. (2000) The costs of diversity: The diversification discount and inefficient investment, *Journal of Finance* **55**, 35–80.

Schenone, C. (2004) The effect of banking relationships on the firm's IPO underpricing, *Journal of Finance* **59**, 2903–2958.

Schmid, M. M. and Walter, I. (2009) Do financial conglomerates create or destroy economic value? *Journal of Financial Intermediation* **18**, 193–216.

Sette, E. and Gobbi, G. (2015) Relationship lending during a financial crisis, *Journal of European Economic Association* **13**, 453–481.

Shin, H. S. (2009) Reflections on Northern Rock: The bank run that heralded the global financial crisis, *Journal of Economic Perspectives* **23**, 101–120.

Shleifer, A. and Vishny, R. W. (2010) Unstable banking, *Journal of Financial Economics* **97**, 306–318.

Stein, J. (1997) Internal capital markets and the competition for corporate resources, *Journal of Finance* **52**, 111–133.

Stein, J. (2002) Information production and capital allocation: Decentralized versus hierarchical firms, *Journal of Finance* **57**, 1891–1921.

Stiroh, K. J. (2004) Diversification in banking: Is noninterest income the answer? *Journal of Money, Credit and Banking* **36**, 853–882.

Sufi, A. (2009) Bank lines of credit in corporate finance: An empirical analysis, *Review of Financial Studies* **22**, 1057–1088.

UBS. (2008) Shareholder report on UBS's write-downs. Available at: http://www.ubs.com/1/ShowMedia/investors/releases?contentId=140331&name=080418ShareholderReport.pdf.

Williamson, O. (1975) Markets and Hierarchies*: Analysis and Antitrust Implications*, Collier Macmillan Publishers, New York.