



HAL
open science

On the Costs and Externalities of Intermediation

Mario Milone

► **To cite this version:**

Mario Milone. On the Costs and Externalities of Intermediation. Business administration. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLED070 . tel-03222723

HAL Id: tel-03222723

<https://theses.hal.science/tel-03222723>

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à **Université Paris Dauphine**

On the Costs and Externalities of Intermediation

École doctorale n°521

SCIENCES DE GESTION

Spécialité FINANCE

Soutenue par **Mario MILONE**

le 3 Décembre 2018

Dirigée par **Gilles CHEMLA**

COMPOSITION DU JURY :

M. Gilles Chemla, Directeur de thèse
Directeur de Recherches CNRS
Université Paris Dauphine

M. Denis Gromb, Rapporteur
Professeur de Finance
HEC Paris

Mme. Kathy Yuan, Rapporteur
Professor of Finance
London School of Economics

M. Christopher Hennessy, Membre du jury
Professor of Finance
London Business School

M. Jérôme Dugast, Membre du jury
Associate Professor of Finance
Université Paris Dauphine

To my mother,

Aknowledgments

I am greatly indebted to my advisor, Professor Gilles Chemla, for his invaluable advice, support, and patience. For all the time and efforts that you have spent and all the knowledge that you have brought me, I deeply thank you.

I would like to express my sincere gratitude to Professor Oliver Hart who has warmly welcomed me in the economics department of Harvard University and who has inspired me at great length.

This thesis would not have been as diverse and rich without the invaluable contribution of Professor Katrin Tinn with whom the third chapter has been coauthored.

I am thankful to *all* the professors I had the pleasure to interact with during this thesis, particularly Professor Christopher Henessy, professors at Université Paris Dauphine, Harvard University, and Imperial College Business School, who have provided helpful advice and comments.

I want to especially thank all the PhD students with whom I have shared the struggles and uncertainties of this PhD with, and who have always been of great support in that journey.

Last but not least, I am immensely grateful to my family, my late mother, my father, my brothers Emmanuel and Simon, my great uncle Antoine, and all my family members who have provided invaluable support and encouragements, and without whom this thesis would not have been possible.

Contents

1	Introduction	3
2	Smart Lending	11
2.1	Introduction	13
2.2	The traditional lender's problem	19
2.2.1	Agents and Technologies	19
2.2.2	Optimal Contract	23
2.2.3	Screening with data	26
2.3	Technological lender and coexistence with the traditional lender	32
2.3.1	Technological lender	32
2.3.2	Coexistence with the traditional lender	33
2.4	Dynamic implications of a data-based screening technology . .	39
2.5	Discussion and conclusion	41
	Bibliography	44
	Appendices	51
2.A	The technological lender's problem	51
2.B	Proofs	55
3	Bank Asset Structure and Risk-Taking	59
3.1	Introduction	61
3.2	Literature	64
3.3	The Model	67
3.3.1	Liquidity shock	73
3.3.2	Portfolio choice	75
3.4	Regulatory Implications	79

3.4.1	Capital Regulation	81
3.4.2	Liquidity Regulations	82
3.4.3	Impact of liquidity shock	83
3.5	Conclusion	84
	Bibliography	86
	Appendices	91
4	Assessing Transit Rents	97
4.1	Introduction	99
4.2	A model of global trade with transit rents	106
4.2.1	The setting	106
4.2.2	Gravity equation and multilateral resistance terms	109
4.2.3	Welfare	111
4.2.4	Monopolistic fees charged	112
4.2.5	Example with four countries	112
4.3	Global Trade Network	119
4.4	Data and empirical strategy	123
4.5	Results	126
4.5.1	Existence of transit rents	126
4.5.2	Global welfare implications	129
4.5.3	Role of FTAs and customs unions	134
4.5.4	Further discussion and robustness	137
4.6	Conclusion	138
	Bibliography	141
	Appendices	153
	Global Bibliography	163

Chapter 1

Introduction

In a world where capital is scarce and unequally distributed among economic agents, financial markets provide an essential service: allocating capital to its best possible use. Optimal capital allocation is arguably the *raison d'être* of the financial sector, and markets, by allowing agents to trade with one another, contribute to the efficient use of capital. However, if markets are efficient, why would there be any use for firms or financial intermediaries?

In his famous article, Coase (1937) provides an answer by bringing forward the importance of transaction costs. Indeed, market-based transactions can be costly due to the lengthy process of agreeing on a price and the negotiation and writing of a contract. Consequently, some transactions may be more efficiently carried out inside an organization like a firm rather than externally through the use of markets. Coase's ideas, originally strongly debated as they represented a complete paradigm shift from the neo-classical theory of the firm, have irreversibly shaped the way economists think about firms and intermediaries. This departure has spawned a vast literature on the understanding of the firm and the determinants of its boundaries that aims at addressing the weaknesses and incompleteness of Coase's theory, namely the fact that he did not specifically address the reasons why transactions might be cheaper inside organizations. Williamson (1971) and Williamson (1975) introduced the idea that transactions may be more efficiently carried out inside firms because *authority* – between an employer and its employee

for instance – provides an effective way to solve problems when unexpected events occur. While leaving open the question of the enforceability of that authority, Williamson's, and later Klein, Crawford, and Alchian (1978)'s contributions represent a significant progress in the understanding of the differences between markets and organizations, such as intermediaries. Grossman and O. Hart (1986) and O. Hart and Moore (1990) provide a better understanding of the forces driving the boundaries of the firm. They argue that property rights determine who makes decisions in the case unspecified events occur, which explains why some agents may prefer to own assets – or be independent contractors – and why some are employees – and do not own the asset.¹ While originally developed for an understanding of the nature of the firms, Coase's theory and the recognition of costly transactions provide a justification for the existence of financial intermediaries, such as banks. The main operations that banks perform consist in granting loans and receiving deposits from the public. In other words, banks and financial intermediaries can be seen as specializing in buying financial securities when granting loans and selling them when accepting deposits. The reason why depositors would not directly be lending their capital via the capital markets (buying bonds for instance) can be explained by the existence of transaction costs. Through the pooling of depositor resources, financial intermediaries can economize on these costs and as such be seen as coalitions of agents benefiting from economies of scope. While providing useful services to the economy, intermediaries can also generate costs and externalities.

In a three chapters journey, this thesis aims at determining and estimating some costs of intermediation and the magnitude of the externalities that intermediaries can generate. The first two chapters focus on financial intermediation while the third one studies intermediation in the context of international trade.

Minimizing transaction costs through economies of scope is not the only service provided by financial intermediaries. In order to optimally allocate depositors capital, banks perform extremely important functions for the econ-

¹For a more in depth presentations of the evolution of the theory of the firm, refer to O. Hart (1989) and O. Hart (2011) or Chemla and Milone (2018).

omy, namely screening and monitoring. These represent a fundamental difference between the way banks and capital markets operate. In financial markets, optimal capital allocation is ensured by the determination of prices, invisibly directing agents' capital to its most productive use. However, when a firm approaches a bank for financing, the bank needs to assess whether investing in that firm represents a good use of depositors' money. The assessment of the creditworthiness of borrowers – screening – is arguably one of the most fundamental service provided by banks to the broader economy. As intermediaries, banks are in the best position to specialize in screening as depositors do not have to duplicate the cost of doing so themselves. Moreover it may be the only possible way for some firms to access financing (e.g. R. Rajan (1992)). Banks have long been viewed as possessing superior abilities to screen borrowers because they typically entertain relationships with their customers, thereby having access to information that no one else has (e.g. R. Rajan (1992), Petersen and R. Rajan (1994), Petersen and R. Rajan (2002a), Karolyi (2018), and Inderst and Mueller (2007)). However, recent years have witnessed the rise in new types of lenders. Those do not rely on long term relationships to provide financing but instead make use of the increased availability of data and new complex predictive algorithms. The existence of these new types of intermediaries – fundamentally similar in the service they provide, but different in the ways they do so – raises several new questions. How do these new entrants differ from existing intermediaries? What are the consequences of their increasing presence on the economy? Do they provide a threat to financial stability? Most importantly, the fundamental question is: how do we understand them? The first chapter of this thesis proposes a framework to understand the impact of these changes by framing the problem as informational. Indeed, big data, machine learning and the rise of Fintech are fundamentally changing the way information is processed and is currently revolutionizing financial intermediation. While the potential benefits of technology, such as speed and cost efficiency may be easy to appreciate, its costs are far less understood. To investigate the impact of technology on financial intermediation, I define and capture the key characteristics in which technology may impact the screening process in

terms of information. I find that while technology can improve information processing, it can also exacerbate financial frictions. Additionally, the existence of new types of intermediaries create externalities on existing ones and lead to an increase in the cost of traditional financial intermediation.

While providing useful services to the economy, financial intermediaries also pose a threat to financial and economic stability. This has been illustrated once more during the last financial crises of 2008. However it is because financial intermediaries are able to be so beneficial to the economy that they can be particularly destructive. This is arguably the main externality of financial intermediaries and the reason why regulators all over the world spend considerable amounts of time trying to ensure that not only banks, but also other financial institutions serve the economy without threatening its stability. Before the financial crisis of 2008, the dominant approach of banking regulation was microprudential. It was based on the belief that if every bank was safe, then it had to be true that the system as a whole was safe as well. Economists were forced to review this approach in the wake of the crisis as they realized the importance of liquidity problems, whereby an agent suddenly face the impossibility to honor claims or repay depositors. This is particularly problematic when a financial intermediary faces such problems as it snowballs across the economy: people who were expecting money from the intermediary themselves face the impossibility to honor their own claims. To cope with that problem, and to ensure that liquidity problems do not become “contagious”, the latest banking regulations of Basel III require banks and other financial intermediaries to hold considerably more liquid assets – assets that can be used to weather bad economic conditions – than in the past in the hope that doing so will mitigate, or even avoid, future crises. However, it may prove difficult for regulators to be sure that banks hold assets that are indeed liquid in case of trouble. For that reason, regulators now engage in so called “stress-testing” exercises where they ask banks to simulate bad economic scenarios in order to verify that they would indeed be able to cope with them. Essentially, regulators want to make sure that if something bad were to happen, not all the assets of the bank would suddenly become worthless. That is, they would like banks to

hold assets that are not too correlated with each others. The second chapter of this thesis aims at providing a framework to understand the incentives for bank to increase or decrease the correlation of their assets. It is because the measurement of this risk is difficult that a theoretical analysis of the banks' incentives is needed in order to guide current and future regulations. To do so I analyze the dual objective of liquidity management and risk management by banks. I find that capital requirements – the main regulatory tool used before the crisis – can have adverse effects by inducing banks to increase the correlation between their assets. However the newly introduced liquidity requirements can mitigate this behavior. This result is important because it highlights a trade-off between these two regulatory tools – with regard to this particular risk –, while intuition dictates that they should both enhance financial stability.

Some intermediaries are not the response of the existence of frictions such as transaction costs, but are de facto intermediaries. Such examples can be found in the context of international trade. International trade, not unlike the financial sector, improves the use of capital by allowing countries to specialize in the production of goods that each of them are the most apt to produce. In doing so, resources are saved and consumption and welfare improved for all parties participating in trading relationships. And not unlike financial markets, trading is subject to frictions, the most prominent one of them being transportation costs, naturally arising from the distance separating trading partners. Another friction, identified as far back as Smith (1776) and Coase (1937), results from the fact that sometimes the trade between two countries has to transit through another third party country. For instance a non coastal country that does not have a direct access to a port, also denoted landlocked, is particularly subject to this problem. “Transit countries” – the ones between two trading partners – are, by many accounts, intermediaries: they are effectively intermediating goods between two trading partners. While their existence is exogenously determined by geographical constraints, they nonetheless provide a needed service by permitting trade. In fact, the problem created by these intermediaries is very close in spirit to the hold-up problem studied in the contract theory literature, whereby the impossibility

to write contracts that can anticipate every future contingencies – incomplete contracts – can create situations where some contracting parties can be placed in undesirable future bargaining positions, introducing inefficiencies in ex-ante choices. Identically, countries that have no other choice than trading through transit countries can be considered “held up”. The last chapter of this thesis aims at shedding light on the cost of these intermediaries – transit countries. This chapter develops a theoretical understanding of this friction and highlights its differences with more well known transportation costs. It also provides empirical evidence of the economic significance of this problem and estimate its global impact on welfare. Interestingly, it is shown that countries that are seemingly unaffected by this friction (e.g. The United Kingdoms or The United States) are in fact negatively affected by it from the distortionary effect it creates on global prices. As such, the existence of these intermediaries and their direct costs are the sources of global externalities.

Bibliography

- Chemla, Gilles and Mario Milone (2018). “Oliver Hart - La Finance Vue à Travers La Théorie Des Contrats Incomplets”. *Les Grands Auteurs En Finance, 2ème Édition*. Éditions EMS.
- Coase, R.H. (1937). “The Nature of the Firm”. *Economica* 4.16, pp. 386–405.
- Grossman, Sanford and Oliver Hart (1986). “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”. *Journal of Political Economy* 94.4, pp. 691–691.
- Hart, Oliver (1989). “An Economist’s Perspective on the Theory of the Firm”. *Columbia Law Review*.
- (2011). “Thinking about the Firm: A Review of Daniel Spulber’s The Theory of the Firm”. *Journal of Economic Literature* 49.1, pp. 101–113.
- Hart, Oliver and John Moore (1990). “Property Rights and the Nature of the Firm”. *Journal of Political Economy* 98.6, pp. 1119–1158.
- Inderst, Roman and Holger M. Mueller (2007). “A Lender-Based Theory of Collateral”. *Journal of Financial Economics* 84.3, pp. 826–859.
- Karolyi, Stephen Adam (Feb. 1, 2018). “Personal Lending Relationships”. *The Journal of Finance* 73.1, pp. 5–49.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978). “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process”. *The Journal of Law & Economics* 21.2, pp. 297–326.
- Petersen, Mitchell A. and Raghuram Rajan (1994). “The Benefits of Lending Relationships: Evidence from Small Business Data”. *The Journal of Finance* 49.1, pp. 3–37.
- (2002a). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *Journal of Finance* 57.6, pp. 2533–2570.

- Rajan, Raghuram (1992). “Insiders and Outsiders: The Choice between Informed and Arm’s-Length Debt”. *The Journal of Finance* 47.4, pp. 1367–1400.
- Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Reprint, Liberty Classics, 1979.
- Williamson, Oliver E. (1971). “The Vertical Integration of Production: Market Failure Considerations”. *The American Economic Review* 61.2, pp. 112–123.
- (1975). *Markets and Hierarchies: Analysis and Antitrust Implications : A Study in the Economics of Internal Organization*. Free Press. 312 pp.

Chapter 2

Smart Lending

This paper shows that a data-based screening technology can increase the cost of financial intermediation. The use of data in the screening process reduces the acquisition of soft information by traditional lenders, which harms constrained borrowers further. Additionally, groups in which fewer borrowers were financed in the past are under-represented in the data, leading to a cross-sectional difference in screening efficiency. Screening is more efficient for borrowers with greater historical lending data. When traditional and technological lenders coexist, the borrowers about whom data can provide precise information raise funds from technological lenders while those with less informative historical data choose traditional lenders who can make up for the lack of hard data-based information by acquiring soft information. The intermediation cost for the traditional lender is increased by the existence of technological lenders. I identify conditions under which traditional lenders benefit from restricting their own access to data-processing technology when competing against the technological lender.

2.1 Introduction

Information processing is entering a new age. Fast increases in the production and storage of data and in computing capacity are the two major ingredients for the big data revolution which quickly transforms the finance industry.¹ Today’s top mortgage originator in the U.S. is Quicken, which uses automated algorithms and underwriting to perform lending decisions.² The rise of Fintech, big data, and machine learning has widely been recognized by the academic community, but *“there is no reason to think [...] that these innovations will automatically enhance stability or even access of service”* (Philippon, 2016).

This paper investigates how the intensive use of data in the lending process impacts financing decisions. I argue that technological and traditional lenders have very different attitudes toward screening, which is perhaps the most important aspect of capital allocation. The ability for financiers to assess the quality of investment opportunities, projects or mortgages, is of critical importance to a well functioning economy, and is the *raison d’être* of financial intermediaries. As an information intensive activity, screening is likely to be hugely impacted by the increasing use of data. Quicken’s website “Rocket Mortgage” advertises a mortgage approval process of only a few minutes, thanks to an automated screening process of applicants. While the benefits of a more automated and intensively data-based screening technology such as speed and cost-efficiency may be easy to appreciate, its costs are far less understood. This paper shows that a data-based screening process is likely to aggravate financial frictions and intensify credit rationing.

Assessing borrowers credit-worthiness is a complex task that requires various sources of information. The literature typically distinguishes between two types: hard information, which is viewed as factual, quantitative, and easy to store and to substantiate, and soft information, which can only be

¹A 2017 report from IDC forecasts a ten-fold increase in data-creation over the next 10 years, from 16 Zetabytes in 2016 to more than 160 in 2025.

²Quicken has replaced Wells Fargo as top US retail mortgage lender in Q1 2018. Among the top ten U.S. mortgage originators, six of them can be considered Fintech firms in the sense that the use of automated processing of data and information is at the center of their business process.

acquired through human interaction and is more subjective and difficult to communicate and/or store (see e.g. Liberti and Petersen (2017), Berg, Puri, and Rocholl (2013), and Agarwal and Ben-David (2014)). Understanding the difference between these two types of information is highly relevant in the context of information processing, as not all types can easily be used by automated algorithms. While the definitions of soft and hard information varies across the literature, this paper defines hard information as being processable by a machine and by predictive algorithms, and soft information as being acquirable only through human interaction.

This paper identifies two channels through which the use of data in the screening process can aggravate financial constraints created by the presence of moral hazard. First, the use of historical lending data based on hard information lowers the incentives for traditional lenders to acquire and process soft information in assessing the creditworthiness of borrowers. As a result, borrowers for whom soft information is an important dimension in determining their creditworthiness become more financially constrained when data is used in the screening process. Second, one of the key characteristics of data-based knowledge is that the precision of that knowledge reflects the availability and structure of the data used to construct it. In the context of lending, a data-based screening technology tends to be more precise and more efficient for applicants that are more represented in the data. Borrowers that are already financially constrained due to the presence of moral hazard become even more so *compared* to the ones that are less affected by moral hazard, as the lender has a better ability to screen the latter than the former. Importantly, this second channel does not result from possible existing biases present in the data. Rather, it is driven by the difference in the precision of the information extracted from the data. Since lending decisions depend on the efficiency of the screening process, the difference in knowledge precision inherent to a data-based approach tends to negatively affect groups of borrowers that are under-represented in existing data. These groups are in turn less likely to obtain credit, they have less data generated about them, which further increases the difference in the precision of screening across groups.

I also analyze an economy in which both traditional and technological

lenders coexist. The traditional lender has both the ability to use historical lending data and to acquire soft information to screen borrowers. In contrast, the technological lender can only use data, but is more efficient at doing so. I show that when both traditional and technological lenders coexist, borrowers tend to separate between the two depending on their individual characteristics. Borrowers whose soft information is important in determining their creditworthiness prefer to seek financing from traditional lenders while the others are siphoned off by technological lenders. Interestingly, borrowers staying with the traditional lender may be better off if she does not make a heavy use of data in her screening process. Additionally this siphoning-off effect increases the intermediation costs of traditional lender because she faces a pool of borrowers whose average screening cost increases due to the entrance of the technological lender.

I examine these effects in a two periods model of firm financing featuring moral hazard and a data-based screening technology. Borrowers are heterogeneous in both their amount of net worth and the extent to which soft information matters in assessing their creditworthiness. In the first period, the traditional lender does not have access to historical lending data and only makes use of soft information to screen borrowers. The second period lending game is affected by the data generated during the first period. The existence of historical lending data affects the incentives for the traditional lender to exert costly effort for acquiring soft information which is detrimental to the borrowers for whom soft information matters. Additionally, the presence of moral hazard in the first period lending results in richer borrowers being more financed than poorer borrowers. As a consequence, in the second period, the lender has access to more historical data about rich borrowers which relaxes their financing constraint compared to poor borrowers. In the second period I allow for a purely technological lender – one who only uses data to screen borrowers – to coexist with the traditional lender. Because the technological lender is more efficient at extracting information from the data, she is able to siphon-off (from the traditional lender) borrowers for whom soft information is relatively less important in determining their creditworthiness. This separation result affects the average screening

cost for the traditional lender. Because she exerts more effort for borrowers whose soft information matters, the entrance of the technological lender in the credit market *increases* the average screening cost of the traditional lender. While the technology efficiencies are assumed exogenous at first, I subsequently study the incentives for the traditional lender to restrict her use of data processing technology. A trade-off is revealed whereby increasing her use of data allows the traditional lender to compete with the technological lender and attract borrowers whose soft information is relatively less important. However, doing so increases credit-rationing for borrowers whose soft information is important. The traditional lender thus prefers to lower her technological efficiency when the population of borrowers contain relatively more of those whose soft information matters.

This paper is first and foremost related to the vast banking literature on screening and information. It has been repeatedly argued that one advantage of bank financing comes from banks ability to acquire and process soft information through lending relationships thereby providing superior screening and monitoring services (e.g., Allen (1990) and D. Diamond (1984)). Boot (2000) defines relationship banking as the process in which the lender gathers information (i) “beyond readily available public information”, (ii) “over time through multiple interactions with the borrower”, and that (iii) “remains confidential (proprietary)”. Petersen and R. Rajan (1994) provide empirical evidence showing that stronger relationships between lenders and borrowers are associated with increased availability of credit for small firms. Petersen and R. Rajan (1995) argue that competition in credit market lowers the value of lending relationships as it becomes harder for lenders to internalize the associated benefits. However, Petersen and R. Rajan (2002b) provide empirical evidence that the increase in information technology allows lenders to provide credit to firms that are more distant, and with whom communication is more impersonal (see also Berger, Frame, and Miller (2005)), and Jeremy C. Stein (2002) argues that it can lead to larger and more centralized banks, which can be detrimental to small-business lending.

The model captures that characteristic of bank lending by giving the traditional lender the unique ability to use soft information in the screening

process. While the dynamic effects of such borrower-lender relationship have been the subject of many studies (e.g. Bolton and David S. Scharfstein (1990), Dewatripont and Maskin (1995), and D. Diamond (1991)), this paper does not consider repeated relationships and the existence of possible long-term contracts between lenders and borrowers. While featuring a dynamic two periods setting, each period is considered as a static problem in the spirit of Innes (1990). Relatedly, I focus on the screening (e.g. Inderst and Mueller (2006)) rather than monitoring (e.g. (Townsend (1979) and Gale and Hellwig (1985)) problem faced by the lender. While contracting in every period is a static problem in our model, the interesting dynamic effects come from the use of a data-based screening technology by the lender.

A strand of literature considers the role of information sharing among lenders. Padilla and Pagano (2000) and Bennardo, Pagano, and Piccolo (2015) find theoretically that information sharing can increase borrowers effort and reduce their incentive to over-borrow. We also have empirical evidence that information sharing lowers default rates (Jappelli and Pagano, 2002; Brown, Jappelli, and Pagano, 2009), increases private credit (Djankov, McLiesh, and Shleifer, 2007) but may also exacerbate credit run (Hertzberg, Liberti, and Paravisini, 2011). These papers consider information sharing as being the exchange of information between lenders of a given borrower. For example Padilla and Pagano (2000) study the incentives of a borrower to exert effort knowing that information about her might be shared with other lenders. In contrast, this paper assumes that borrowers only live one period and that data relates to historical information about past borrowers that can be used to help assessing the creditworthiness of new borrowers that have never been lent to before.

A more recent literature has focused on the impact of credit-scoring and data on lending decisions and monitoring. There is evidence that the use of transaction accounts can help monitoring (Mester, Nakamura, and Renault, 2007; Norden and Weber, 2010) as well as screening (Puri, Rocholl, and Steffen, 2017). Berg, Burg, et al. (2018) show that digital information left by potential borrowers via the use of digital services – denoted digital footprint – provide relevant information for the purpose of assessing the creditworthiness

of borrowers. Relatedly, Jagtiani and Lemieux (2018) provide evidence that the use of alternative data sources improve the assessment of creditworthiness by providing different type of information compared to the traditional FICO scores. Relatedly, Paravisini and Schoar (2015) and Berg, Puri, and Rocholl (2016) study how the use of credit score and hard information affect lenders incentives to exert effort. Our model encompasses these findings by allowing for a substitutability between the use of data and the use of soft information.

Understanding the impact of information and data has become increasingly relevant in recent years with the rise in data availability and predictive algorithms. More specifically, Fuster et al. (2018) study the impact of the use of non-linear prediction techniques and their impact on mortgage lending and show that an increase in technology complexity may lead to discrimination in the mortgage lending market. Our paper relates to Fuster et al. (2018) by capturing an essential characteristic of data-based screening technology, specifically the fact that the structure of the data affects the performance of the assessment of borrowers creditworthiness in their cross-section.

This paper considers the problem as being informational in nature and relates to the role of information in investment decisions (e.g. Cabrales, Gossner, and Serrano (2013), Meyer (1991), Meyer and Zwiebel (2007), and Shorrer (2015)), the role of information in the principal-agent relationships (e.g. Levitt and Snyder (1997) and Chaigneau, Edmans, and Gottlieb (2017)), and more importantly to the literature on learning and dynamic inattention (e.g. Ellis (2018), Mayskaya (2017), Matejka and McKay (2015), and Nimark and Sundaresan (2018)). In fact, one way to appreciate the problem of using hard and/or soft information from the lender's point of view is to consider them as two sources of information that the lender can choose from and focus more or less attention on, as in Van Nieuwerburgh and Veldkamp (2010) and Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016). U. Rajan, Seru, and Vig (2015) provide evidence that algorithms used in the assessment of default risk for the securitization of subprime mortgages in the period 1997-2006 failed to take into account important characteristics relevant for credit-risk, generating worse pools of loans over time.

The rest of the paper proceeds as follows. I first present the problem of

the first period lending by the traditional lender where only soft information is used in the screening process. I then consider the second period lending in which the traditional lender has the ability to use historical lending data and I show the effect on the borrowers financing constraint. I then introduce the purely technological lender and study the coexistence of the two types of lenders. I show how borrowers separate between the lenders and that the average screening cost of the traditional lender increases. I then turn to the choice of technological efficiency for the traditional lender and identify conditions under which she prefers to restrict her use of data. I subsequently study the dynamics effect of the use of data on financial constraints and describe the dispersion effect – where the dynamic interaction of lending and screening increases the effect of moral hazard. I finally discuss the results and conclude.

2.2 The traditional lender's problem

This section studies the problem of a traditional lender. The traditional lender has the ability to use both soft information and historical lending data, when available, to screen borrowers. I first focus on the traditional lender's problem in the first period, where historical data is not available, and only soft information is used to screen borrowers. I then study the second period lending decision where the traditional lender can use the first period lending data to help in the screening process. In the next section, a purely data-based lender, denoted technological lender will be introduced and the coexistence of both types of lenders will be studied.

2.2.1 Agents and Technologies

There is a representative lender and good and bad borrowers in equal proportion. Each borrower, indexed by k , has access to a project requiring an investment of 1. A good borrower's project returns R with probability q if the borrower behaves. If the borrower shirks, the project returns zero, but the borrower enjoys a private benefit B_k . In case of failure, the project returns

zero. A bad borrower's project returns zero for sure, regardless of whether the borrower behaves or shirk. The returns for good and bad borrowers are illustrated in figure 2.1. The borrower's type – good or bad – is unobservable to *both* the representative lender and the borrower and we assume that good borrowers that behaves have positive NPV projects.

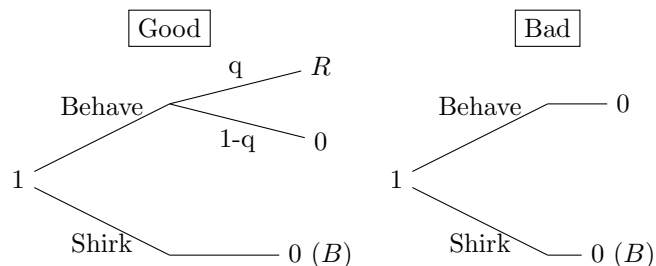


Figure 2.1: Projects returns

Assumption 1. *Good borrowers have positive NPV projects when they behave*

$$qR > 1$$

Assuming that both the lender and the borrower are unaware of the borrower's type ensures symmetry of information at all point in time. While strong, this assumption effectively removes adverse selection problems arising from information asymmetries. This allows to focus on the impact of moral hazard alone.

There are two periods $t = \{0, 1\}$. At the beginning of each period, N borrowers are born and live for one period, hence dismissing any possibility for long-term relationship or contracts. Each borrower is endowed with net worth $A_k \in (0, 1)$ and private benefit B , unique for all borrowers. Additionally, each borrower k is characterized by a parameter γ_k that captures how important soft information matters in determining his credit-worthiness. The role of γ_k will become clear shortly.

A_k is uniformly distributed over $(0, 1)$ and γ_k is distributed on $(0, \bar{\gamma})$ following the distribution $\Gamma(\cdot)$. Finally, each borrower is associated with a vector of individual characteristics X_k that are predictive of his type. Net

worth A_k , private benefit B , importance of soft information γ_k and characteristics X_k are perfectly observable by all agents.

Every period, borrowers apply to the lender for credit, who decides whether or not to finance each of them. Each borrower approaches the lender and asks for the lender to finance $(1 - A_i)$ of the project's required investment. The borrower may choose to use all of his net worth, in which case $A_i = A_k$ or only part of it and $A_i < A_k$. The lender then screens the borrower. The screening outcome is modeled by an informative signal about the borrower's type, observed by the lender. Upon receiving the signal, the lender updates her belief about the success probability of the borrower and offers a contract specifying the repayment r_k to the lender if the project succeeds. The lender receives a binary signal $s_{t,k} \in \{0, 1\}$ with precision $\tau_{t,k} \in \{\frac{1}{2}, 1\}$ such that

$$P[s_{t,k} = 1|Good] = P[s_{t,k} = 0|Bad] = \tau_{t,k}$$

As good and bad borrowers exist in equal proportion in the economy, and because borrowers do not observe their types, the unconditional probabilities that a good or bad borrower approaches the lender for financing are equal. Additionally, and without loss of generality, it is assumed that screening is necessary for financing in that the lender needs to acquire information before accepting to finance the borrower.

Assumption 2. *The prior belief about the borrower's type is such that*

$$P[Good] = P[Bad] = \frac{1}{2}$$

and screening is necessary for financing

$$qR < 2$$

Assumption 2 ensures that the lender refuses to finance borrower k unless she has access to an informative signal about his type. As a result, financing is always refused if the lender receives a signal $s_{t,k} = 0$. In the first period (period 0), the traditional lender does not have access to historical lending data, and the signal precision only depends on how much soft information is

acquired and processed. The first period signal precision follows the following functional form.

$$\tau_{0,k} = \frac{1}{2} + \gamma_k e_k \quad (2.1)$$

e_k denotes the effort exerted by the lender to acquire soft information. In the context of this model, soft information refers to the type of information that can be acquired through human interaction with the borrower. As all borrowers are born in the beginning of the period, it does not include information acquired over time (through long term relationship). γ_k captures the importance of soft information in assessing the credit-worthiness of borrower k . When soft information matters more (higher γ_k), acquiring soft information is more valuable and the signal precision is higher, given a level of effort. Acquiring and processing soft information is costly for the lender in the following way.

$$c(e_k) = \alpha \left(\frac{e_k^2}{2} \right) \quad (2.2)$$

For simplicity and tractability, I assume that the cost of effort is a quadratic function of the effort exerted by the lender. The results go through as long as the cost of effort is an increasing and convex function. The parameter α is a scaling factor ensuring that at the optimum level of the lender's effort, the signal precision is not greater than 1.

The timing is as follow, and illustrated in figure 2.2. At the beginning of the period, borrower k approaches the lender and ask her to finance $(1 - A_i)$ of the project financing cost. A_i is a choice variable for the borrower such that $A_i < A_k$ where A_k is the total wealth of borrower k . Then, the lender decides to exerts effort e_k at cost $c(e_k)$. After exerting the effort, the lender receives the signal $s_{0,k}$ with precision $\tau_{0,k}$. Upon receiving the signal, and if the lender decides to finance the project, she offers a contract r_k that maximizes her expected profits as will become clear shortly. Subsequently, if the borrower accepts the contract, he decides to behave or shirk, and finally returns realize and payoffs are distributed. The next subsection solves the optimal contract.

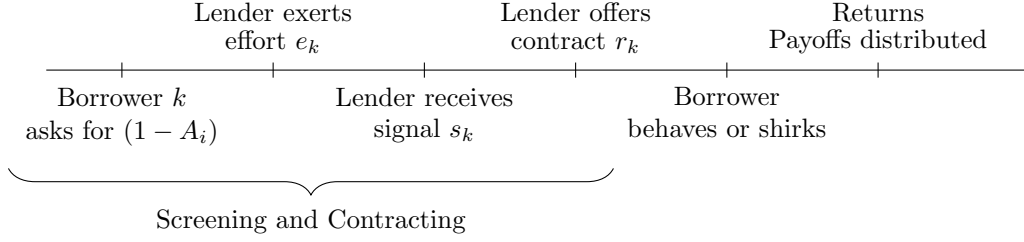


Figure 2.2: One period contracting timeline

2.2.2 Optimal Contract

The optimal contract is solved by backward induction. From assumption 2, the lender only consider financing borrower k if she receives a signal $s_{0,k} = 1$, and the ex-ante probability to receive such a signal is $\frac{1}{2}$, regardless of effort e_k . Let us denote $p(e_k)$ the posterior belief about the probability of success of the borrower conditional on receiving $s_{0,k} = 1$ and that the borrower behaves. The borrower solves the following program.

$$\begin{aligned}
 & \max_{A_i} \frac{1}{2} [p(e_{eq})(R - r_k) - A_i] + A \\
 \text{s.t. } & p(e_{eq})(R - r_k) \geq B & (IC_B) \\
 & e_{eq} = \arg \max_e \frac{1}{2} [p(e_k)r_k - (1 - A_i)] - c(e_k) & (IC_L) \\
 & p(e_{eq})(R - r_k) \geq A_i & (PC_B) \\
 & \frac{1}{2} [p(e_{eq})r_k - (1 - A_i)] - c(e_{eq}) \geq 0 & (PC_L) \\
 & A_i < A & (LL_A) \\
 & r_k < R & (LL_r)
 \end{aligned}$$

The borrower chooses A_i to maximize his expected profit given effort e_{eq} exerted by the lender. Given A_i , the lender chooses the the level of effort e_{eq} that maximizes her profits minus the cost of effort, as described by the incentive compatibility constraint of the lender, equation (IC_L) . If the lender receives the signal $s_{0,k} = 1$, she offers the contract r_k where r_k is the repayment to the lender if the project succeeds. r_k is such that the borrower

is incentivized to behave, and needs to satisfy the incentive compatibility constraint of the borrower, inequation (IC_B). Equations (PC_B) and (PC_L) ensure that both parties want to participate, and equations (LL_A) and (LL_r) are limited liability constraints.

It is important to note that the incentive compatibility constraint of the borrower, inequation (IC_B), is a function of $p(e_{eq})$, the posterior belief of the lender after receiving the signal $s_{0,k} = 1$. This is due to the fact that ex-ante, neither the lender nor the borrower know whether the borrower is good or bad. However, after the lender offers the contract r_k , the borrower can update his belief about his own type as r_k is a sufficient statistic. His decision to behave or shirk is therefore based on the same updated belief than the lender. As the lender offers the contract, she chooses r_k such that the borrower receives just enough to behave.

$$r_k = R - \frac{B}{p(e_{eq})}$$

The program can be rewritten (omitting the limited liability constraints for readability) as follows.

$$\max_{A_i} \frac{1}{2} [B - A_i] + A_k$$

$$\text{s.t. } e_{eq} = \arg \max_e \frac{1}{2} \{ [p(e_k) R - 1] - [B - A_i] \} - c(e_k) \quad (2.3)$$

$$\frac{1}{2} [p(e_{eq}) R - 1] - c(e_{eq}) \geq \frac{1}{2} [B - A_i] \geq 0 \quad (2.4)$$

Inequation (2.4) combines the participation constraints of the lender and the borrower. Note that if $B < A_i$, the borrower is better off keeping his assets rather than investing them into the project while if A_i is too low the lender chooses not to participate. Hence, the borrower wants to choose A_i sufficiently low so that the constraint (PC_B) is satisfied and sufficiently high so that constraint (PC_L) is also satisfied. Given A_i , the lender chooses effort $e_{eq}(A_i)$ that satisfies the first order condition of equation (2.3). The borrower in turn chooses A_i to maximize his profits given the effort $e_{eq}(A_i)$ that the lender will exert. Lemma 1 lays down the optimal contract. It is important

to note that while the lender exerts an effort that maximizes her profit, the borrower chooses A_i such that the lender is in zero-profits.

Lemma 1. Optimal contract without data

The optimal contract is such that the lender's effort satisfies

$$e_{eq} = \frac{1}{2\alpha} qR\gamma_k \quad (2.5)$$

and the borrower's profits when the lender does not use historical lending data are

$$\pi_k = \frac{1}{2} \left[\frac{1}{2} qR - 1 \right] + \frac{1}{2\alpha} \left(\frac{1}{2} qR\gamma_k \right)^2 \quad (2.6)$$

Lemma 1 highlights that the borrower's profits are increasing in the importance of soft information for assessing the borrower's creditworthiness. As γ_k increases, the lender exerts higher level of effort as it is more valuable to acquire soft information. In turn, screening is more efficient and the borrower's expected profits are higher. Plugging the optimal effort e_{eq} into the participation constraint (2.4) allows us to find the condition for financing to occur.

$$\pi_k(\gamma_k) \geq \frac{1}{2} (B - A_k) \quad (2.7)$$

Equation (2.7) makes clear that borrower k needs to finance enough of the project in order to obtain financing. If the borrower is not able to finance enough of the project's cost, it is too expensive for the lender to incentivize him to behave, and financing does not take place. This result is common in models of financing in the presence of moral hazard. Note that $\pi_k(\gamma_k)$ is increasing in γ_k . Therefore, for a given level of net worth A_k , borrower k needs to have a γ_k high enough so that it is profitable for the lender to exert effort in the screening process. On aggregate, only the fraction of borrowers with enough wealth A_k and/or high enough γ_k are able to access financing, due to the presence of moral hazard. Figure 2.3 illustrates the financing constraint resulting from moral hazard.

The horizontal axis of figure 2.3 is the importance of soft information in assessing the creditworthiness of borrower k , γ_k . The increasing convex curve

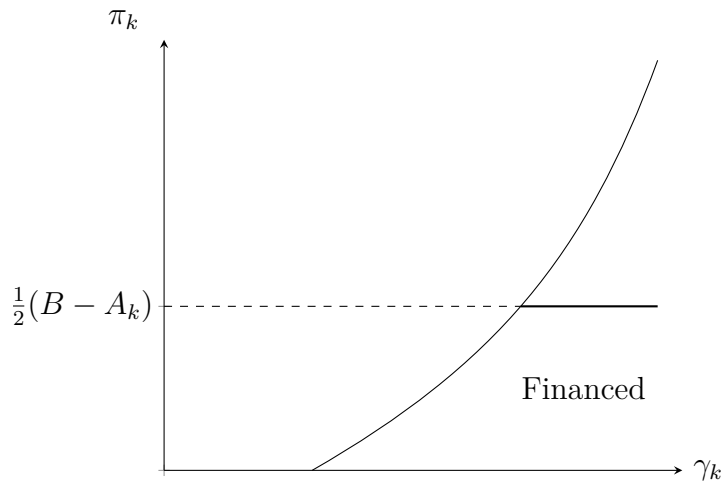


Figure 2.3: Financial constraint

is the borrower's profits π_k under the optimal contract. The horizontal line represents the right hand side of the financial constraint (2.7). The higher the net worth A_k of borrower k , the lower the horizontal line. Financial constraint (2.7) states that all borrowers that are under the curve are able to obtain financing, as their combinations of (A_k, γ_k) satisfy the financial constraint arising from the presence of moral hazard. The solid horizontal line on the right of the curve can be interpreted as the number of borrowers with a given net worth that are able to access financing. As the net worth of the borrower decreases, fewer borrowers are able to access financing (only those for whom soft information is informative enough). The slope of the curve represents the marginal effect of moral hazard on the financial constraint, as it captures how much more a borrower is constraint when his wealth marginally decreases.

We now turn to the second period lending to study how the usage of data in the screening process impacts the financial constraint of borrowers.

2.2.3 Screening with data

At the beginning of the second period, the lender has access to historical lending data consisting of all the borrowers that have been financed in the first period. The lender has lent to borrowers with observable characteristics

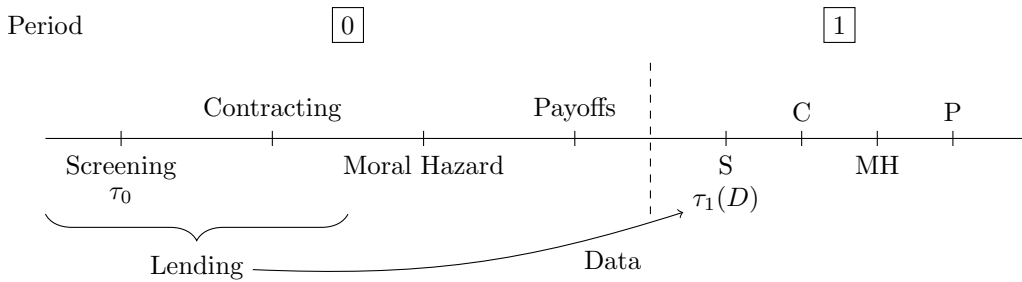


Figure 2.4: Timeline

X_k and has also observed whether those borrowers have defaulted or not, let us denote this information the output Y_k . As has been described previously, the observable characteristics X_k are predictive of the borrower type, good or bad. Using the period one data (X_k, Y_k) the lender can try to learn the mapping between X_k and Y_k in order to predict better the creditworthiness of the borrowers applying for credit in period 2. The way this is performed in practice is through the use of predictive algorithms. These could be logistic regressions or more complex algorithms from the field of machine learning. Such a learning problem, where both the inputs X_k and outcomes Y_k are used to learn the mapping between the two is typically tackled with the use of so called supervised predictive algorithms. For a more detailed explanation of the learning process, the reader is referred to section 2.A of the appendix. Figure 3.1 illustrates the dynamic link between the first period lending decisions and the second period screening.

The model does not seek to capture the details of the algorithm used by the lender to infer the mapping between X_k and Y_k , as many of them are available and they are constantly evolving. Instead, it seeks to capture fundamental characteristics of such a learning process in order to understand the implications for lending decisions. I assume that the outcome of learning using historical lending data depends on two factors. First, the lender can learn better the mapping between X_k and Y_k if more data is available. Second, the learning is of better quality if the technology used – performance of the algorithm – is more efficient. In other words, the lender can extract more information from historical lending data if (i) there is more data available

and/or (ii) if the technology used is better able to extract useful information from the available data. It is important to note that the information given by past lending is not dependent on having the same borrowers seeking financing. Instead, it is assumed that all borrowers from the first period die before the beginning of the second period, and that newly born borrowers ask for financing in the second period. Nevertheless, the functional dependence between X_k and Y_k is similar enough between the first and the second period such that historical lending data is informative when screening new borrowers in the second period. Historical lending data is best thought of as capturing hard information that is useful in assessing the creditworthiness of the borrower.

Let us denote D the amount of data available to learn the mapping between X_k and Y_k and η the efficiency of the technology in use. The use of historical lending data modifies the precision of the second period $\tau_{1,k}$ and the cost of effort $c(e_k)$ as follows.

$$\tau_{1,k} = \frac{1}{2} + \gamma_k e_k + \eta D \quad (2.8)$$

$$c(e_k) = \alpha \left(\frac{e_k^2}{2} + \eta D e_k \right) \quad (2.9)$$

Equations (2.8) and (2.9) capture two characteristics of the link between hard and soft information. First, it captures the fact that hard and soft information provide different types of information useful in assessing the borrower's creditworthiness. That is, even if one has access to hard information, it might still be useful to exert effort to acquire soft information. Second, equation (2.9) captures the substitutability between hard and soft information. If the lender already has access to historical lending data to assess the borrower's credit worthiness, the effort required to increase the precision of the screening signal will be higher because both sources of information aim at reducing the same uncertainty. The informational problem can equivalently be seen as having the lender accessing two sources of information, hard and soft. Both sources reduce the uncertainty about the borrower's type. If the lender already has access (costlessly) to the hard information source, the

marginal value of the soft information lowers – equivalently, the marginal cost of soft information increases. The program solved by the borrower is identical to the previous case, except for the change of functional forms. The optimal contract and effort is laid out in the next lemma.

Lemma 2. Optimal contract with data

The optimal contract is such that the lender's effort satisfies

$$e_{eq,D} = \frac{1}{2\alpha} qR\gamma_k - \eta D \quad (2.10)$$

and the borrower's profits when the lender uses historical lending data D are

$$\pi_{k,D} = \pi_k + \frac{1}{2}\eta D [qR(1 - \gamma_k) + \alpha\eta D] \quad (2.11)$$

Financing is possible if and only if

$$\pi_{k,D}(\gamma_k, D) \geq \frac{1}{2}(B - A_k) \quad (2.12)$$

The new optimum effort exerted by the lender is now lowered by the fact that historical lending data is available to help the lender in screening the borrowers. When more data is available, the marginal benefit of soft information is lower – equivalently, the marginal cost of soft information is higher –, and the lender optimally reduces the acquisition of soft information. Equation (2.11) shows that the borrower's profits when the lender can use data equals to the profits when no data is used plus an additional term increasing in the amount of data and technology efficiency ηD and decreasing in the the importance of soft information γ_k . On the one hand having more data available to screen borrower k increases the signal precision and the borrower's profits. On the other hand, more data lowers the use of soft information by the lender, which negatively impacts borrowers for whom soft information is important in determining their creditworthiness. When γ_k is high enough, the fact that the lender has access to historical data is detrimental to the borrower, compared to the case where the lender does

not have access to data. In other words, the availability of data discourages the lender's effort to use soft information, which negatively affects borrowers with high γ_k .

Lemma 3. *There exists $\tilde{\gamma}$ such that*

- *If $\gamma_k < \tilde{\gamma}$, the use of historical lending data increases the expected profits of borrower k .*
- *If $\gamma_k > \tilde{\gamma}$, the use of historical lending data decreases the expected profits of borrower k .*

where $\tilde{\gamma} = 1 + \alpha \frac{\eta D}{qR}$.

An important implication of the use of data by the traditional lender is that it increases the difference in financial constraints between richer and poorer borrowers, effectively increasing the impact of moral hazard on credit-rationing. To see that, one need to look at the slope of $\pi_{k,D}$ – with data – and how it differs from the one of π_k – without data –, because the financial constraint takes an identical form in both cases, as shown by inequation (2.12). The slope of the borrowers profits indicates the extend to which borrowers with low net worth are constrained *compared* to borrowers with high net worth. It can easily be shown that the slope of $\pi_{k,D}$ is lower than the one of π_k indicating that when the lender uses data when screening borrowers, there is a larger difference in the financial constraint of borrowers depending on their net worth. This leads to the next proposition.

Proposition 1. *The availability of historical lending data in the screening process accentuates the effect of moral hazard on credit-rationing.*

Figure 2.5 illustrates this result. Subfigure 2.5a shows the difference in credit-constraint between borrowers with different net worth. The lower the threshold on the horizontal axis, the higher the borrower's net worth A_k (as the threshold is $\frac{1}{2}(B - A_k)$). There are more borrowers with higher net worth that can access financing as the minimum γ_k that is required is lower. When the traditional lender uses data in the screening process (subfigure 2.5b), the

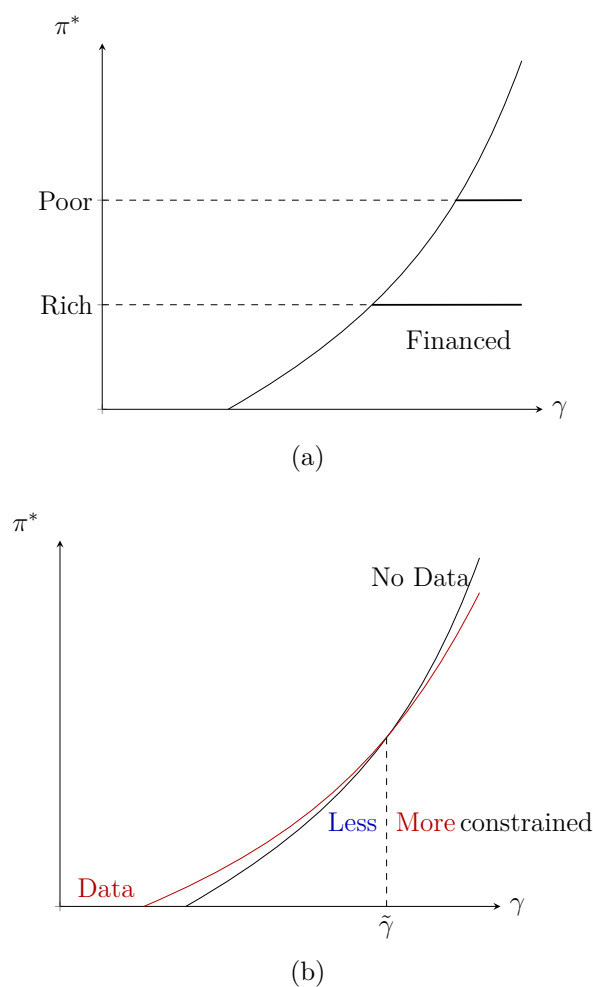


Figure 2.5: Effect of the use of historical lending data on credit-constraint.

slope of the borrowers profits lowers, and the profits are equal at the point $\tilde{\gamma}$. As the slope decreases, the difference between the credit constraints of rich and poor borrowers increases.

In the next section, I introduce a purely technological lender and subsequently study the outcome of the coexistence of both types of lenders.

2.3 Technological lender and coexistence with the traditional lender

2.3.1 Technological lender

We observe a rise in the presence of lenders whose screening process are mainly – if not only – based on the use of data. One can think of Quicken loans in the mortgage market. Most of these lenders only offers an online application process and promise much faster processing of the borrowers applications than that of traditional lenders. The main difference with traditional lenders is that these “technological lenders”, as they are referred to in this paper, do not interact with the borrowers personally. Instead they make an extensive use of historical lending data to predict the quality of the applying borrowers in order to decide whether credit should be granted. In a way, these lenders make a greater use of hard information of borrowers, the type of information that can easily be transmitted through electronic systems and verified. That being said, the model acknowledges the fact that this hard information can be used to partly proxy what might be characterized as being soft information. For instance, assume that the lender somehow has access to the shopping pattern of the borrower applying for credit. That might inform the lender about the type of shopper the borrower is – cautious or not for instance. This substitutability between hard and soft information is captured in the functional form used in the traditional lender’s problem.

I define a technological lender as one that can only make use of hard information, at zero cost. I additionally assume that the technological lender is better able to extract information from the available data. One may consider that the technological lender is specializing in the use of predictive technologies and has an advantage on that area compared to the traditional lender, perhaps through ex-ante investment in the technology. Once the technology is ready to use, however, one can imagine that the technological lender only has to “push a button” to assess the creditworthiness of the borrower. While the difference of technological efficiencies between the traditional and technological lenders is assumed at first, I later consider the choice of the

2.3. TECHNOLOGICAL LENDER AND COEXISTENCE WITH THE TRADITIONAL LENDER³

traditional lender in her use of data and identify conditions under which she might want to reduce her reliance on such technologies.

The precision of the signal received by the technological lender is

$$\tau_1^T = \frac{1}{2} + \eta^T D$$

where $\eta^T > \eta$ is the technological efficiency of the technological lender.

The optimal contract and the program solved by the borrower is identical as with the traditional lender albeit simpler as the technological lender does not have the ability to exert any effort. The signal precision only depends on the amount of historical data available, D , and η^T . Moral hazard is still present thereby financing is still constrained.

Lemma 4. *Borrower k can access financing through the technological lender if and only if*

$$\pi^T = \frac{1}{2} [qR\tau_1 - 1] \geq \frac{1}{2} (B - A_k)$$

Lemma 4 makes clear that the financial constraint of the technological lender is now independent of the importance of soft information, γ_k , in assessing the borrower's credit worthiness³. Figure 2.6 plots the financial constraint for the technological lender. Similarly to the case of the traditional lender, the borrower can only access financing if he can cover a large enough part of the project's cost. On the figure, all borrowers with a sufficiently large net worth (borrowers below the line π^T) can access financing.

2.3.2 Coexistence with the traditional lender

One of the question of interest when studying the impact of the use of data is to analyze how the entrance of a purely technological lender affects the traditional lender. To that end, I introduce both types of lenders and ask

³Note that the fact that the financial constraint of the technological lender is independent of γ_k is not robust to the functional form used. One may intuitively expect that the borrower's profits might decrease in the important of soft information as the technological lender does not have the ability to use it. The subsequent results would go through even in that case however.



Figure 2.6: Financial constraint for the technological lender

three questions. First, I study the choice of borrowers between asking financing to either of the lender. Second I show how the cost of intermediation for the traditional lender is affected by the entrance of the technological lender. Finally, I study the choice of the technological efficiency of the traditional lender under the assumption that both lenders compete for market share in the credit market.

Borrowers separation

When both types of lenders are available, the borrowers can choose between seeking financing to either of them. The choice of lender depends on the expected profits of the borrowers. As we already have derived the borrowers profits for both lenders in the previous sections, we can readily analyze the borrowers choices. Figure 2.7 illustrates the borrowers profits for both lenders.

As $\pi_{k,D}$ is increasing in γ_k , π^T is constant in γ_k , and $\eta^T > \eta$, there exists a threshold γ_{sep} such that $\pi_{k,D}(\gamma_{sep}) = \pi^T$. Depending on the important of soft information in assessing his creditworthiness γ_k , borrower k prefers to be financed by either the technological or traditional lender. This result is formally stated in the next proposition.

2.3. TECHNOLOGICAL LENDER AND COEXISTENCE WITH THE TRADITIONAL LENDER

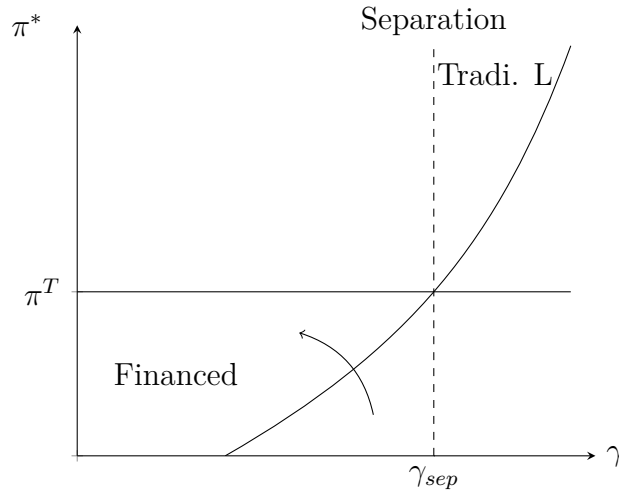


Figure 2.7: Borrowers choice of lender

Proposition 2. *When both traditional and technological lenders coexist, there exists a threshold γ_{sep} such that*

- *If $\gamma_k < \gamma_{sep}$, borrower k prefers to seek financing from the technological lender.*
- *If $\gamma_k > \gamma_{sep}$, borrower k prefers to seek financing from the traditional lender.*

Proposition 2 has a very intuitive interpretation. The traditional lender has a comparative advantage over the technological lender as she can use soft information to assess the borrowers creditworthiness. This ability is valuable for borrowers whose soft information matters as it increases their expected profits. Therefore high γ_k borrowers prefer to seek financing from the traditional lender as her ability to discriminate them is higher. On the other hand borrowers whose soft information is not very important in determining their credit worthiness prefer to seek financing from the technological lender. Because the technological lender has a higher technological efficiency and because soft and hard information are substitutable to some extent, borrowers with lower γ_k are better off asking for financing from the technological lender. The entrance of the technological lender on the credit-market

creates a *siphoning-off effect* on the traditional lender pools of borrowers, as illustrated by the curved arrow of figure (2.7).

Cost of traditional intermediation

Proposition 2 has consequences in terms of screening costs for the traditional lender. The optimal effort exerted by the traditional lender increases in γ_k since the marginal value of acquiring soft information is larger when γ_k is higher. As the screening cost is increasing in effort, it is costlier (but optimal) for the traditional lender to screen applicants with higher γ_k . Proposition 2 states that only borrowers with high γ_k stay with the traditional lender while the lower γ_k borrowers that used to be financed by the traditional lender shift away to the technological lender. As a result, the pool of borrowers faced by the traditional lender after the entrance of the technological lender exhibits a higher average γ_k . This leads to the next proposition.

Proposition 3. *The entrance of the technological lender increases the average screening cost of the traditional lender.*

It is interesting to interpret this result in light of Philippon (2016)'s paper. Philippon (2016) attempts to assess the potential impact of FinTech on the finance industry and documents the fact that while the financial sector has seen an entrance of new type of financial intermediaries, financial services remain expensive. While he argues that the cost of financial services can explain the entrance of new types of intermediaries, the result of proposition 3 suggests that the entrance of new intermediaries exhibiting lower costs may result in the increase in the cost of more traditional intermediaries, due to the fact that both lenders serve different types of borrowers, and the traditional lender is left with costly borrowers.

Choice of technological efficiency for the traditional lender

One assumption that has been made insofar is that the technological lender somehow has a superior ability to extract information from available historical lending data. In this section, I study the endogenous choice of the lender to modify her technological efficiency.

2.3. TECHNOLOGICAL LENDER AND COEXISTENCE WITH THE TRADITIONAL LENDER 3

To do so, one needs to define the objective function of the lenders. In the current setup, both lenders are in perfect competition and earn zero-profit in expectation for each loan that is extended to borrowers. That section assumes that the lenders compete for market share in the credit market. One way to capture these incentives is to allow for the lenders to charge a fee ε for each borrower that receives financing. Note that ε can be thought of as being as small as desired, in line with the position of high competition between lenders. It turns out that the traditional lender faces a trade-off in her choice of technological efficiency η , as illustrated in figure 2.8.

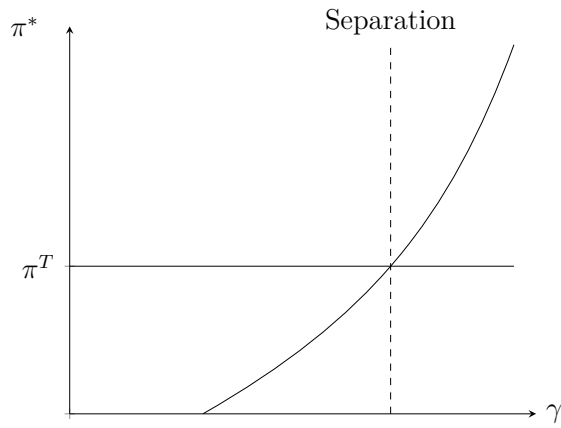
Lemma (3) provides the basis for this trade-off. On the one hand, when the traditional lender increases her technology efficiency, she is able to attract borrowers for whom soft information is relatively less important. As shown in the subfigure 2.8b, increasing η shifts the separation threshold of the borrowers to the left, indicating that the traditional lender is able to capture a larger fraction of borrowers with average γ_k . On the other hand, increasing η results in more severe credit-rationing for the high γ_k borrowers. As seen on the figure, borrowers that lie between the red and the black curve for high values of γ_k do not have access to financing any longer if the traditional lender increases η .

The resulting incentives for the traditional lender to increase or decrease her technological efficiency when competing with a technological lender depends on the distribution of the borrowers population along γ_k . Recall that this distribution is denoted Γ . This leads to the next result.

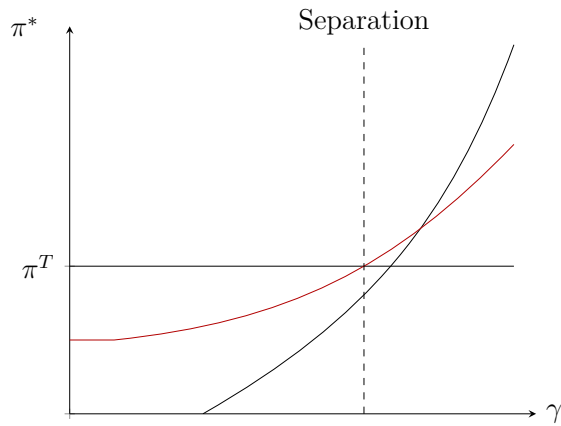
Lemma 5. *The incentives for the traditional lender to choose her technological efficiency is as follows.*

- *If Γ is skewed towards high values of γ_k , the traditional lender prefers to lower her technological efficiency η .*
- *If Γ is skewed towards low values of γ_k , the traditional lender prefers to increase her technological efficiency η .*

Lemma (5) characterizes the choice of the technological efficiency of the traditional lender as a function of the borrowers distribution along the γ_k



(a)



(b)

Figure 2.8: Effect of an increase in the technological efficiency (η) of the traditional lender.

dimension. In the case where there are many borrowers for whom soft information is important in determining their creditworthiness, it is more valuable for the traditional lender to decrease her technological efficiency because it decreases credit rationing for these borrowers thereby increasing the amount of borrowers to whom credit can be granted. Otherwise, if most borrowers do not rely much on soft information to be assessed, it becomes more profitable for the traditional lender to compete with the technological lender for these borrowers, and increase her technological efficiency.

2.4 Dynamic implications of a data-based screening technology

In this section, I analyze further the dynamic impact of the use of a data-based screening technology. In the previous sections, I have analyzed the case in which the lender has access to an amount of historical data D informative of the borrower's type. It was then implicitly assumed that the same amount of information was available to the lender regardless of the borrower being assessed. However, the credit constraint of equation (2.7) makes clear that the presence of moral hazard constraints borrowers differently along their net worth. As illustrated in subfigure 2.5a, at the end of the first period lending, the lender gathers more data about rich borrowers compared to poor borrowers. As a result, the second period lending suffers from an imbalance in the structure of the data used in the screening process. This section aims to show that this dynamic effect between lending decision and data structure results in an increase of the effect of moral hazard. This effect is denoted the *dispersion effect*.

Let us first consider, as was previously assumed, that the lender has access to the same amount of informative data regarding borrowers of any net worth. In that case, the financing constraint is the one defined by equation (2.12) and the borrower's profits satisfy equation (2.11). The black curve of figure 2.9 plots the financial constraint in that case.

Let us now consider the fact that due to moral hazard, the first period lending results in an imbalanced historical lending data where richer borrowers are more represented than poorer borrowers. In that case, the lender's optimal effort result in the following borrowers profits.

$$\pi_{k,D} = \pi_k + \frac{1}{2}\eta D(A_k) [qR(1 - \gamma_k) + \alpha\eta D(A_k)]$$

where $D(A_k)$ captures the amount of informative historical data as a function of the borrower's net worth A_k .

The presence of moral hazard in the first period lending creates a positive relationship between the borrower's net worth and the amount of informative

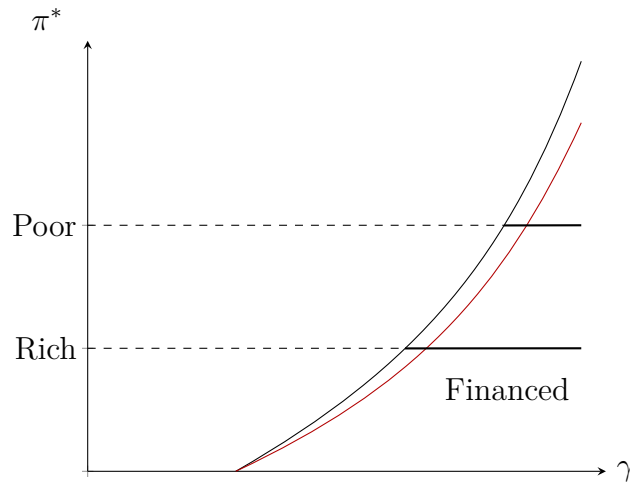


Figure 2.9: Dispersion effect

data available in the second period lending:

$$\frac{\partial D(A_k)}{A_k} > 0$$

The implication of that relationship is that poorer borrowers are now at a disadvantage compared to richer borrowers in the sense that the lender has a superior ability to screen richer borrowers using their hard information. As a consequence, compared to the case where the lender has the same amount of informative data to screen any borrower, the profit of a poor borrower is necessarily lower. Therefore, poor borrowers, already constrained from the presence of moral hazard, become even more so *compared* to richer borrower because the lender has *relatively* less informative data to assess their creditworthiness.

The red curve of figure 2.9 illustrates the impact of the data structure resulting from the first period lending on the financial constraint and leads to the result of the next proposition.

Proposition 4. *Dispersion Effect*

The credit constraint resulting from the presence of moral hazard is increased by the use of data in the screening process through the imbalance in historical lending data moral hazard creates.

It is important to note that some conditions need to hold for the dispersion effect to arise. Recall that assessing the creditworthiness of the borrowers using historical lending data is performed through the use of predictive algorithm whereby the lender aims to find the mapping between the borrowers observable characteristics X_k and the outcome (success/default) Y_k . As is explained in more details in section 2.A of the appendix, predictive algorithms are less efficient at assessing the creditworthiness of under-represented groups of borrowers under the assumption that the mapping between X_k and Y_k is not identical for all borrowers. In other words, it is necessary that the functional form (or the data generating process) f_k such that $Y_k = f_k(X_k)$ is not identical for all borrowers.

2.5 Discussion and conclusion

The novelty of this paper is to investigate the effects of the use of data in the screening process, both through its interaction with the use of soft-information as well as dynamically. I first show that the use of data in the screening process lowers the lender's incentives to acquire costly soft information, which is detrimental to borrowers for whom soft information is important in determining their credit-worthiness. While the functional forms used in the current model may seem restrictive, they aim to capture important characteristics of the interaction between hard and soft information. Namely, they captures the fact the hard information can be a substitute to soft information, and that when lenders can use data, it becomes harder to improve the screening efficiency using soft information. Such results are related to the inattention literature, and one can view the lender as having a choice between using two sources of information, one denoted hard and another denoted soft. The fact that both sources are substitutable to some extent is important for the results of the paper. In fact, one can imagine that if both hard and soft information were to be complement, it might be that the traditional lender would always have a superior screening ability, ruling out the presence of purely technological lender, something we in fact observe.

I also show that the use of data in the screening process creates a dynamic effect arising from the interaction between lending and screening, denote the dispersion effect. Data-based technologies are fundamentally dependent on the structure of available observations which creates an inter-dependence between the lending decisions and the screening process. Such inter-dependence may lead to a screening technology that exhibits dispersion in its efficiency across borrowers. Namely, borrowers that are under-represented in the data are at a disadvantage in the screening process compared to other borrowers. It is important to note that the presence of moral hazard provides a fundamental endogenous reason as to why the data might be imbalanced, resulting in historical data where financially constrained borrowers become under-represented. This last channel is novel in that it does not rely on the fact that the data-based technology might exhibit biases (as in Fuster et al. (2018)), but arises because the predictions performances of such algorithms depend on the data that is provided to them. In that regard, it is close but different than a discrimination story whereby some individuals would be penalized because the technology provides biased estimates, non-fundamentally justified by their differences.

In addition to shedding light on two possible channels through which the use of data may increase financial frictions, I find that when some lenders specialize in the use of data for screening borrowers (technological lender), the structure of the lending market changes in a way that affects the cost of intermediation of traditional lenders. This result hinges on the siphoning-off effect technological lenders have, leading only a certain type of borrowers to shift away from traditional lenders. Borrowers separate between the traditional and technological lender in such a way that the traditional lender is left with borrowers associated with higher screening costs – borrowers for whom soft information matters. This results in an increase of average screening costs borne by the traditional lender. While this result is not dependent on the dispersion effect of the data-based technology, it is strengthened when the technology exhibits such a characteristic.

Finally, I show that when a traditional lender competes with a technological lender, she may have incentives to limit the use of her data-based

screening technology and specialize in her core competency that is the use of soft information. This is due to the fact that the traditional lender faces a trade-off whereby increasing its use of data allows to capture borrowers serviced by the technological lender, but also rations credit for borrowers whose soft information matters. Therefore, while the model assumes that traditional lenders are endowed with a less efficient technology, it provides reasons to believe that this can arise endogenously.

Overall the results of this paper are in line with the finding of Philippon (2016) showing that while we observe a rise in the presence of new types of lenders, the cost of intermediation does not seem to decrease, as would be expected by a typical competition argument. Instead, my model suggests that competition might incentivize traditional lenders to specialize in their soft information screening technology, even though it increases the average cost of intermediation.

Bibliography

- Agarwal, Sumit and Itzhak Ben-David (Feb. 2014). *Loan Prospecting and the Loss of Soft Information*. Working Paper 19945. National Bureau of Economic Research.
- Allen, Franklin (Mar. 1, 1990). “The Market for Information and the Origin of Financial Intermediation”. *Journal of Financial Intermediation* 1.1, pp. 3–30.
- Bennardo, Alberto, Marco Pagano, and Salvatore Piccolo (Mar. 1, 2015). “Multiple Bank Lending, Creditor Rights, and Information Sharing”. *Review of Finance* 19.2, pp. 519–570.
- Berg, Tobias, Valentin Burg, et al. (Apr. 2018). *On the Rise of FinTechs - Credit Scoring Using Digital Footprints*. Working Paper 24551. National Bureau of Economic Research.
- Berg, Tobias, Manju Puri, and Jörg Rocholl (May 2013). *Loan Officer Incentives and the Limits of Hard Information*. Working Paper 19051. National Bureau of Economic Research.
- (Oct. 25, 2016). *Loan Officer Incentives, Internal Rating Models and Default Rates*. Working Paper ID 2022972. Rochester, NY: Social Science Research Network.
- Berger, Allen N., W. Scott Frame, and Nathan H. Miller (2005). “Credit Scoring and the Availability, Price, and Risk of Small Business Credit”. *Journal of Money, Credit and Banking* 37.2, pp. 191–222. JSTOR: 3838924.
- Bolton, Patrick and David S. Scharfstein (Mar. 1, 1990). “A Theory of Predation Based on Agency Problems in Financial Contracting”. *The American Economic Review* 80.1. ArticleType: research-article / Full publication

- date: Mar., 1990 / Copyright © 1990 American Economic Association, pp. 93–106. JSTOR: 2006736.
- Boot, Arnoud W. A. (Jan. 1, 2000). “Relationship Banking: What Do We Know?” *Journal of Financial Intermediation* 9.1, pp. 7–25.
- Brown, Martin, Tullio Jappelli, and Marco Pagano (Apr. 1, 2009). “Information Sharing and Credit: Firm-Level Evidence from Transition Countries”. *Journal of Financial Intermediation* 18.2, pp. 151–172.
- Cabrales, Antonio, Olivier Gossner, and Roberto Serrano (2013). “Entropy and the Value of Information for Investors”. *American Economic Review* 103.1, pp. 360–377.
- Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (Aug. 1, 2017). *Does Improved Information Improve Incentives?* SSRN Scholarly Paper ID 2269380. Rochester, NY: Social Science Research Network.
- Dewatripont, M. and E. Maskin (1995). “Credit and Efficiency in Centralized and Decentralized Economies”. *The Review of Economic Studies* 62.4, pp. 541–555.
- Diamond, Douglas (1984). “Financial Intermediation and Delegated Monitoring”. *The Review of Economic Studies* 51.3, pp. 393–414. JSTOR: 2297430.
- (1991). “Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt”. *Journal of Political Economy* 99.4, pp. 689–721.
- Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer (May 1, 2007). “Private Credit in 129 Countries”. *Journal of Financial Economics* 84.2, pp. 299–329.
- Ellis, Andrew (Jan. 1, 2018). “Foundations for Optimal Inattention”. *Journal of Economic Theory* 173, pp. 56–94.
- Fuster, Andreas et al. (July 26, 2018). *Predictably Unequal? The Effects of Machine Learning on Credit Markets*. SSRN Scholarly Paper ID 3072038. Rochester, NY: Social Science Research Network.
- Gale, Douglas and Martin Hellwig (1985). “Incentive-Compatible Debt Contracts: The One-Period Problem”. *The Review of Economic Studies* 52.4, pp. 647–663. JSTOR: 2297737.

- Hertzberg, Andrew, José María Liberti, and Daniel Paravisini (Apr. 1, 2011). “Public Information and Coordination: Evidence from a Credit Registry Expansion”. *The Journal of Finance* 66.2, pp. 379–412.
- Inderst, Roman and Holger M. Mueller (2006). “Informed Lending and Security Design”. *Journal of Finance* 61.5, pp. 2137–2162.
- Innes, Robert D (Oct. 1, 1990). “Limited Liability and Incentive Contracting with Ex-Ante Action Choices”. *Journal of Economic Theory* 52.1, pp. 45–67.
- Jagtiani, Julapa and Catharine Lemieux (Apr. 5, 2018). *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the Lendingclub Consumer Platform*. SSRN Scholarly Paper ID 3178461. Rochester, NY: Social Science Research Network.
- Jappelli, Tullio and Marco Pagano (Oct. 1, 2002). “Information Sharing, Lending and Defaults: Cross-Country Evidence”. *Journal of Banking & Finance* 26.10, pp. 2017–2045.
- Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp (Mar. 1, 2016). “A Rational Theory of Mutual Funds’ Attention Allocation”. *Econometrica* 84.2, pp. 571–626.
- Levitt, Steven D. and Christopher M. Snyder (1997). “Is No News Bad News? Information Transmission and the Role of “Early Warning” in the Principal-Agent Model”. *The RAND Journal of Economics* 28.4, pp. 641–661. JSTOR: 2555780.
- Liberti, José María and Mitchell A. Petersen (Mar. 2017). *Information: Hard and Soft*. Working Paper.
- Matejka, Filip and Alisdair McKay (Jan. 2015). “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model”. *American Economic Review* 105.1, pp. 272–298.
- Mayskaya, Tatiana (Aug. 4, 2017). *Dynamic Choice of Information Sources*. SSRN Scholarly Paper ID 2863605. Rochester, NY: Social Science Research Network.
- Mester, Loretta J., Leonard I. Nakamura, and Micheline Renault (May 1, 2007). “Transactions Accounts and Loan Monitoring”. *The Review of Financial Studies* 20.3, pp. 529–556.

- Meyer, Margaret A. (1991). “Learning from Coarse Information: Biased Contests and Career Profiles”. *The Review of Economic Studies* 58.1, pp. 15-41. JSTOR: 2298043.
- Meyer, Margaret A. and Jeffrey Zwiebel (Mar. 2007). *Learning and Self-Reinforcing Behavior*. Working Paper.
- Nimark, Kristoffer P. and Savitar Sundaresan (2018). *Inattention and Belief Polarization*. Working Paper.
- Norden, Lars and Martin Weber (Oct. 1, 2010). “Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers”. *The Review of Financial Studies* 23.10, pp. 3665–3699.
- Padilla, A. Jorge and Marco Pagano (Dec. 1, 2000). “Sharing Default Information as a Borrower Discipline Device”. *European Economic Review* 44.10, pp. 1951–1980.
- Paravisini, Daniel and Antoinette Schoar (2015). *The Incentive Effect of Scores: Randomized Evidence from Credit Committees*. Working Paper 19303. National Bureau of Economic Research.
- Petersen, Mitchell A. and Raghuram Rajan (1994). “The Benefits of Lending Relationships: Evidence from Small Business Data”. *The Journal of Finance* 49.1, pp. 3–37.
- (May 1, 1995). “The Effect of Credit Market Competition on Lending Relationships”. *The Quarterly Journal of Economics* 110.2, pp. 407–443.
- (Dec. 1, 2002b). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *The Journal of Finance* 57.6, pp. 2533–2570.
- Philippon, Thomas (Aug. 2016). *The FinTech Opportunity*. Working Paper 22476. National Bureau of Economic Research.
- Puri, Manju, Jörg Rocholl, and Sascha Steffen (July 1, 2017). “What Do a Million Observations Have to Say about Loan Defaults? Opening the Black Box of Relationships”. *Journal of Financial Intermediation* 31, pp. 1–15.
- Rajan, Uday, Amit Seru, and Vikrant Vig (Feb. 1, 2015). “The Failure of Models That Predict Failure: Distance, Incentives, and Defaults”. *Journal of Financial Economics* 115.2, pp. 237–260.

- Shorrer, Ran I. (Dec. 2015). *Entropy and the Value of Informatio for Investors: The Prior-Free Implications*. Working Paper.
- Stein, Jeremy C. (2002). “Information Production and Capital Allocation: Decentralized versus Hierarchical Firms”. *The Journal of Finance* 57.5, pp. 1891–1921.
- Townsend, Robert M (Oct. 1, 1979). “Optimal Contracts and Competitive Markets with Costly State Verification”. *Journal of Economic Theory* 21.2, pp. 265–293.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010). “Information Acquisition and Under-Diversification”. *The Review of Economic Studies* 77.2, pp. 779–805.

Appendix

2.A The technological lender's problem

This section describes the statistical problem faced by a lender, and lays out results specific to the lender's learning problem driving for the subsequent economic findings. The goal of the technological lender is to use past lending data in order to assess the creditworthiness of future borrowers applying for credit. The general form of this learning problem is to approximate a data generation process for prediction purposes and can be stated as follows. Consider a data generating process mapping covariates x into a value y such that

$$y = f(x) + \varepsilon$$

where ε is a random variable with mean zero and variance σ^2 . In the context of lending, the covariates x refer to all the observable characteristics of the borrower applying for credit and y is whether or not the borrower will be successful in the project he is undertaking or not. To learn the function f , the lender has access to historical observations about different borrowers that have been lent to in the past. Her goal is to find a function \hat{f} that approximates the true f using this historical data. If she is able to find an \hat{f} that is close to the true f , the lender is able to efficiently screen any *new* borrower applying for credit. I stress the fact that the lender is not learning about an unobservable state regarding a specific borrower through multiple interactions with that borrower over time, but is instead trying to find a general mapping between any borrower that would be true for any

borrower, even if it is the very first time that borrower applies for credit. That being said, it does not prevent the mapping to depend on the history of the borrower, as the covariates x can include information such as past credit card records.

To learn the mapping f , the lender needs a way, an algorithm, also called a “learner”, to find an approximation \hat{f} . Typically, one defines a loss function $L(\hat{y}, y)$ that captures the cost of predicting $\hat{y} = \hat{f}(x)$ when the true value is y . Because y is a non-deterministic function of x , the learner’s objective is to find the *optimal prediction model*, which is the unique function f^* that minimizes the expected loss $E_y[L(\hat{y}, y)]$ for *every possible* value x . The subscript y denotes that the expectation is taken with respect to all possible values of y given x .⁴ To find f^* , the learner has access to a set d of covariates X and values Y , typically called a training set, corresponding in our setting of the past lending data. Given d , a learner produces a model f_d which maps every x into a value $f_d(x) = y_d$. Because every training set d is different, the *same* learner typically produces different prediction models f_d for each of them. A “good” learner is able to generate a prediction model whose predicted values are *close* to the optimal predictions *regardless* of the set d used to train it. In other other words, a good predictive model is one that performs well *out-of-sample*. Regardless of the sample used to train it, a good learner should generate a model that provides accurate predictions on different samples.

A learner might fail at finding such a model for two main reasons. First, it might generate models whose predictions differ depending on the training set d . We say that the learner exhibits *variance*. Second, it might generate models whose average predictions over multiple training sets d are far from the optimal prediction. We say that the learner exhibits *bias*. It turns out that it is difficult for a learner to exhibit both low variance and low bias, as variance increases and bias decreases in the model complexity. A model that predicts for every x the average of all y in a training set d has low variance and high bias, as the predictions do not vary greatly across training sets while

⁴This definition is independent of the distribution of ε , as the expectation takes into account the probability of y , so it remains valid even if, say, ε is skewed.

being far from the true values. A model that predict for every x the value y associated with the closest observation in the training d has high variance and low bias, as the predictions are very dependent on the training set, but close to the true value when averaged across several training sets.⁵ The main challenge of machine learning is to devise learners that optimally balance this bias-variance trade-off.⁶ Two common techniques to solve this problem are regularization and resampling. Regularization aims at penalizing the use of complex functional forms so that simpler ones are preferred even if they do not minimize the expected loss in-sample. This is performed by directly adding a penalization term within the objective function. In resampling techniques – such as bootstrapping or bagging – several training sets are randomly sampled within the full training data and the final model uses the average predictions.

The intuition behind the economic findings of the paper is to consider how an imbalance in the training set can affect the performances of a statistical learner. To analyze this possibility, consider a training set d with N observations, and two partitions (d_1, d_2) of this training set with (N_1, N_2) observations. We denote (X_1, X_2) the sets of covariates and (Y_1, Y_2) their corresponding values. (x_1, x_2) and (y_1, y_2) are elements of (X_1, X_2) and (Y_1, Y_2) . We allow the two partitions to be generated by different data-generating process f_1 and f_2 . The statistical learner, unaware of any subgroup, is trying to find the function f_d mapping x and y that minimizes the expectation of a given loss function L on the full training set d . The learner searches in the set of all possible functions $\{\hat{f}\}$ and solves

⁵One might therefore wonder why machine learning algorithms do not just generate a large number of models with high variance and low bias using multiple training sets, and use the average prediction as the final model, ideally generating a model very close to the optimal one with low variance and low bias. This idea is the basis of many machine learning methods such as bootstrapping and bagging in order to reduce variance while preserving low bias.

⁶It is worth noting the bias-variance problem vanishes when one considers that a learner has access to an infinite amount of data, as a one can use an infinite amount of training sample to find the optimal model exhibiting low variance and low bias. Similarly, the bias-variance trade-off becomes irrelevant for parametric methods as the functional form determines the amount of bias and variance the learner exhibits.

$$f_d = \arg \min_{\{f\}} E_d [L(\hat{f}(x), y)] \quad (2.13)$$

Because the expectation operator is linear, equation (2.13) reads

$$f_d = \arg \min_{\hat{f}} \left\{ \frac{N_1}{N} E_{d_1} [L(\hat{f}(x_1), y_1)] + \frac{N_2}{N} E_{d_2} [L(\hat{f}(x_2), y_2)] \right\} \quad (2.14)$$

Equation (2.14) makes clear that minimizing the expected loss function on the whole sample is equivalent to minimizing the average of the expected loss functions on each subgroup, *weighted by their proportional sample size*. This observation leads to the result of Lemma 6.

Lemma 6. *Consider a training set d of constant size N containing two partitions (d_1, d_2) of sizes (N_1, N_2) following data generating processes (f_1, f_2) .*

If $f_1 \neq f_2$, the predictive performance of a learner increases on the first group and decreases on the second group as $\frac{N_1}{N_2}$ increases.

Lemma 6 has several intuitive interpretations. First, in terms of information. If one subsample increases in size, the learner has relatively more information about that subsample and generates a model that performs better on that subgroup. Second, equation (2.14) can be understood as follows. A learner minimizing the total expected loss does a better job by minimizing the expected loss on the larger group rather than the one on the smaller group as the former carries a larger weight in the objective function. Lemma 6 states that these intuitions are correct if the two groups are generated by different data generating processes. Intuitively, the results are correct if the information available on one group is not fully informative for predictions on the second group. If both groups follow exactly the same data generating processes, that would not be true.

A very important point is that this lemma applies to *non-parametric* methods such as machine learning techniques that are recently getting widely adopted. Parametric techniques are not subject to the bias-variance trade off described above as they use a fixed model complexity. As instance, an OLS regression find only the best *linear* fit of the data. While both methods

are aiming at minimizing a loss function, the problems are fundamentally different as a non-parametric method has the ability to modify the complexity of the model used to fit the data, while a parametric method does not.

In the rest of the paper, we will assume that the conditions specified in Lemma 6 are satisfied on the problem considered. While it may seem arbitrary to assume so, one needs only to assume that the problem faced by the lender – assess the creditworthiness of borrowers – is sufficiently complex and that the probability that all borrowers are affected in the exact same way by fundamentals is sufficiently small. In fact such an assumption is perfectly in line with the observed rise in the use of machine learning techniques as they are specifically designed to find solutions in a complex world where the statistician is not able or prefers not to specify functional forms, leaving that task to the learning algorithm.

2.B Proofs

Lemma 1

First, note that $p(e_k)$ is the probability of success of the borrower conditional on receiving the signal $s_{0,k} = 1$ and conditional on the borrower behaving. Using Bayesian updating, we have that

$$p(e_k) = q\tau_0(e_k) = q\left(\frac{1}{2} + \gamma_k e_k\right)$$

Given A_i , the lender exerts effort e_{eq} that satisfy the following first order condition (where subscripts denote derivatives)

$$\frac{1}{2}p_e(e_{eq})R = C_e(e_{eq})$$

Using the functional forms, we have

$$e_{eq} = \frac{1}{2\alpha}qR\gamma_k$$

The borrower chooses the optimal A_i^* that maximizes his profits given

that the lender will exert the effort derived above

$$\begin{aligned} & \max_{A_i} \frac{1}{2} [B - A_i] + A_k \\ \text{s.t. } & e_{eq} = \frac{1}{2} qR\gamma_k \\ & \frac{1}{2} [p(e_{eq})R - 1] - c(e_{eq}) \geq \frac{1}{2} [B - A_i] \geq 0 \end{aligned}$$

It is clear that the borrower chooses the maximum value of A_i such that

$$\frac{1}{2} [B - A_i^*] = \frac{1}{2} [p(e_{eq})R - 1] - c(e_{eq})$$

The borrower's profit when financed by the traditional lender is therefore

$$\pi_k = \frac{1}{2} [p(e_{eq}\tau_{1,k}(e_{eq}))R - 1] - c(e_{eq})$$

Replacing $p(e_{eq})$ and $c(e_{eq})$ with the lender's effort derived above, we obtain

$$\begin{aligned} \pi_k &= \frac{1}{2} [qR\tau_{0,k}(e_{eq}) - 1] - \frac{e_{eq}^2}{2} \\ &= \frac{1}{2} \left[\frac{1}{2} qR - 1 \right] + \frac{1}{\alpha} \left(\frac{1}{2} qR\gamma_k \right)^2 - \frac{1}{2\alpha} \left(\frac{1}{2} qR\gamma_k \right)^2 \\ &= \frac{1}{2} \left[\frac{1}{2} qR - 1 \right] + \frac{1}{2\alpha} \left(\frac{1}{2} qR\gamma_k \right)^2 \end{aligned}$$

π_k is strictly increasing and convex in γ_k as

$$\frac{\partial \pi_k}{\partial \gamma_k} = \frac{1}{\alpha} \left(\frac{1}{2} qR \right)^2 \gamma_k = \frac{1}{2\alpha} qR e_{eq}$$

Lemma 2

When the traditional lender has access to historical lending data, the borrower's problem stays identical, only the functional forms change. The effort exerted by the lender satisfy

$$\frac{1}{2} p_e(e_{eq})R = C_e(e_{eq})$$

which gives

$$\begin{aligned}\frac{1}{2}qR\gamma_k &= \alpha(2e_{eq} + \eta D) \\ e_{eq} &= \frac{1}{2\alpha}qR\gamma_k - \eta D\end{aligned}$$

And the borrower's profits are

$$\begin{aligned}\pi_{k,D} &= \frac{1}{2} [p(e_{eq})R - 1] - c(e_{eq}) \\ &= \frac{1}{2} [qR\tau_{1,k}(e_{eq}) - 1] - \alpha \left(\frac{e_{eq}^2}{2} + \eta D e_{eq} \right) \\ &= \frac{1}{2} \left[qR \left(\frac{1}{2} + \gamma_k e_{eq} + \eta D \right) - 1 \right] - \alpha \left(\frac{e_{eq}^2}{2} + \eta D e_{eq} \right) \\ &= \frac{1}{2} \left[\frac{1}{2}qR - 1 \right] + \frac{1}{2}qR \left(\gamma_k \left(\frac{1}{2\alpha}qR\gamma_k - \eta D \right) + \eta D \right) - \alpha \left(\frac{\left(\frac{1}{2\alpha}qR\gamma_k - \eta D \right)^2}{2} + \eta D \left(\frac{1}{2\alpha}qR\gamma_k - \eta D \right) \right) \\ &= \frac{1}{2} \left[\frac{1}{2}qR - 1 \right] + \frac{1}{\alpha} \left(\frac{1}{2}qR\gamma_k \right)^2 - \frac{1}{2}qR\gamma_k\eta D + \frac{1}{2}qR\eta D - \frac{1}{2\alpha} \left(\frac{1}{2}qR\gamma_k \right)^2 - \frac{\alpha}{2}(\eta D)^2 + \frac{1}{2}qR\gamma_k\eta D \\ &= \frac{1}{2} \left[\frac{1}{2}qR - 1 \right] + \frac{1}{2\alpha} \left(\frac{1}{2}qR\gamma_k \right)^2 + \frac{1}{2}qR\eta D (1 - \gamma_k) + \frac{\alpha}{2}(\eta D)^2 \\ &= \pi_k + \frac{1}{2}\eta D [qR(1 - \gamma_k) + \alpha\eta D]\end{aligned}$$

As in the case without data, π_k is increasing and convex in γ_k as

$$\frac{\partial \pi_k}{\partial \gamma_k} = \frac{1}{\alpha} \left(\frac{1}{2}qR \right)^2 \gamma_k - \frac{1}{2}qR\eta D = \frac{1}{2}qR e_{eq}$$

However, the slope is lower when the lender uses data as effort decreases in ηD .

Lemma 3

The borrower's profits when the lender does not have access to data is π_k and the profits when the lender does not have access to data is $\pi_k + \frac{1}{2}\eta D [qR(1 - \gamma_k) + \alpha\eta D]$. The borrower has higher expected profits if the

lender does not have access to data if

$$\begin{aligned} qR(1 - \gamma_k) + \alpha\eta D &< 0 \\ qR(\gamma_k - 1) &> \alpha\eta D \\ \gamma_k &> 1 + \alpha \frac{\eta D}{qR} \equiv \tilde{\gamma} \end{aligned}$$

Note that the optimal signal precision when the lender uses data is

$$\tau_{1,k} = \frac{1}{2} + \frac{1}{2\alpha} qR \gamma_k^2 + \eta D (1 - \gamma_k)$$

at $\gamma_k = 1$, the signal precision is $\frac{1}{2} + \frac{1}{2\alpha} qR$ which is lower than 1 if α is high enough. Also, at $\gamma_k = \tilde{\gamma}$, the precision is $\frac{1}{2} + \frac{1}{2\alpha} qR - \frac{\alpha}{2} \frac{(\eta D)^2}{qR}$.

Chapter 3

Bank asset structure and the
risk-taking implications of
capital and liquidity
requirements

In addition to risky loans, banks hold risky securities that provide uncertain future liquidity. This leads them to choose an asset structure with their desired correlation between liquidity and long term asset returns. We show that liquidity management and risk management concerns lead to a trade-off that creates an inverse relationship between security holdings and aggregate asset risk. Capital requirements mitigate liquidity risk in all future states of the world, thereby reducing the cost of liquidity risk and leading banks to increase aggregate asset risk. Liquidity requirements such as the Liquidity Coverage Ratio (LCR) affect high liquidity shock states and mitigate aggregate asset risk-taking. These results highlight the tension between capital and liquidity regulations in addressing the risk taking incentives of financial intermediaries.

In the 1950s liquid assets were typically 30 percent of British clearing banks' total assets, and these largely consisted of Treasury Bills and cash. At June 2007, cash is about 1/2 per cent and traditional liquidity about 1 per cent of total liabilities.

- Tim Congdon, *Financial Times* - 2007

3.1 Introduction

The academic literature on banking regulation has traditionally focused on the liability side of bank balance sheets while theories related to banking assets have generally focused on the origination, monitoring, and sale of loans as well as portfolio choice.¹ The regulatory implications of the staggering transformation of banking assets over the past five decades have been relatively overlooked: While the percentage of safe, liquid assets remained fairly stable, risky securities have become an important part of banking assets. At the end of 2012, more than 50% of the securities held by US commercial bank consisted of risky securities like MBS, CMO, CMBS, corporate and municipal bonds, and other ABS (Hanson et al. (2015)). Do these risky security holdings represent a threat to financial stability?

This paper examines how this transformation of bank asset structure has affected risk taking and discusses the interplay between bank asset allocation and risk taking, on the one hand, and capital requirements and the liquidity coverage ratio, on the other hand. In contrast to traditional banking models that examine liquidity risk, we allow banks to allocate their assets optimally between risky loans and risky securities. Most existing models assume that bank securities are perfectly safe, akin to cash or perfectly safe government bonds. Introducing risky securities gives rise to a specific type of risk that has been neglected in the literature: the correlation between illiquid and liquid assets. As this correlation represents a risk on the overall portfolios of loans

¹See, among many others, D. W. Diamond and Dybvig (1983), D. Diamond (1984), Gorton and Pennacchi (1995), Cerasi and Daltung (2000), Freixas and Rochet (2008), Parlour and Plantin (2008), Acharya and Schnabl (2009), Hartman-Glaser, Piskorski, and Tchisty (2012), Chemla and Hennessy (2014)

(illiquid assets) and securities (liquid assets), we will refer to it as *aggregate asset risk*. Interestingly, it can also be viewed as a “liquidity” *wrong-way risk* in that it captures the risk that the liquidity of securities deteriorates at the same time as the value – or creditworthiness – of the loans. As in the traditional definition of wrong-way risk – the risk that the credit exposure of counterparty A to counterparty B increases at the same time that the creditworthiness of counterparty B deteriorates – this risk can be specific or general. Specific aggregate asset risk can arise because security returns are fundamentally correlated with loan returns. For instance, a bank is exposed to specific aggregate asset risk if it holds mortgage loans along with mortgage backed securities. General aggregate asset risk can come from macroeconomic factors that affect both the creditworthiness of loans and securities liquidity, such as market freezes events observed during the 2007-2008 financial crisis.²

Interestingly, regulators do appear to recognize aggregate asset risk in their definition of high quality liquid assets. In the latest framework of bank supervision, known as Basel III, the Basel Committee introduced a new set of liquidity regulations with two main new ratios: the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR). The ratio of interest in this study, the LCR, is defined as the amount of High Quality Liquid Assets (HQLA) over the total net cash outflow over the next 30 days. Among the four fundamental characteristics that make an asset a HQLA, one is that it needs to have a “low correlation with risky assets” (Basel Committee (2013), p.13). However, while the existence of a correlation between liquidity and long term asset returns appears to be recognized, the Basel III framework is unclear on how the combined use of capital or liquidity requirements affects bank risk taking. Instead, it is generally claimed that risk is taken care of during the process of what is called “stress testing”, where several predefined scenarios should reveal worrying correlations in bank balance sheets. This paper formally analyses this risk and sheds some lights on how capital and liquidity requirements affect the incentives of banks to take on aggregate asset risk. We show that capital requirements reduce aggregate asset risk while liquidity requirements reduce aggregate risk-shifting.

²See Gorton (2010).

In the model, a bank can invest in long term assets – or loans – and securities. Loans can be thought of as corporate loans that cannot easily be securitized and are fundamentally illiquid because they cannot be liquidated without a loss at the time of a liquidity need. Securities are risky in that they provide uncertain future liquidity. This assumption can be viewed from different angles. Securities can be seen as perfectly liquid but providing uncertain future returns. It is the case of risky asset backed securities in a perfectly rational market where investors have full information about future securities returns. Uncertain future liquidity can also capture time varying liquidity. A specific characteristic of liquidity is that it is not only asset dependent but also time dependent. Securities that are liquid in good times can suddenly become illiquid due to market conditions such as investors sentiments. Regardless of the interpretation, it seems fairly reasonable to assume that securities do not always provide liquidity in every possible future states of the world. This is perhaps even more relevant considering today’s debate about the scarcity of safe assets, as underlined by the IMF in its Global Financial Stability Report (2012, chapter 3). The bank’s endogenous choice of aggregate asset risk is subject to two conflicting forces stemming from liquidity management and risk management motives. The bank is subject to a stochastic liquidity shock on its liabilities due to uncertain withdrawals from depositors, as in D. W. Diamond and Dybvig (1983). Liquidity management aims at limiting the risk of liquidity shortage at the time of deposit withdrawals. Because the bank is subject to limited liability in the worst state of the world, it has incentives to correlate loan returns with securities liquidity in order to maximize future expected returns. Liquidity risk creates aggregate risk-shifting. On the other hand the bank engages in risk-management and wants to minimize returns volatility, as in a traditional portfolio analysis. This limits the correlation induced by liquidity risk and gives rise to an equilibrium choice of aggregate asset risk. Risk management arises from the cost associated with liquidity shortages and captured by fire sales. Because the model introduces risky securities that can act as a liquidity buffer, it provides the opportunity to study interactions between liquidity management and risk management, something that cannot be analyzed with a

traditional portfolio framework – focusing on risk management – or liquidity models – focusing on liquidity risk.

I show that capital and liquidity requirements have two opposite effects on bank incentives because they have different effects in different future states of the world. Both regulations limit liquidity shortages but they differ in their contingency. The effects of capital requirements are state independent and affect all future states of the world. It thus reduces risk management concerns and incentivizes the bank to increase aggregate asset risk. On the other hand, liquidity requirements provide state dependent liquidity thus decreasing the effects of limited liability on risk-loving incentives. Similarly, by imposing a minimum amount of securities to be held by the bank, liquidity requirements provide liquidity in the states where securities are the most liquid. Therefore their impact on the bank is similar to an increase in correlation. Hence, the resulting choice of aggregate asset risk decreases. In other words, liquidity requirements provide desirable liquidity characteristics that would have otherwise been created by an increase in aggregate asset risk. The state independent characteristic of capital requirements fails to capture this effect.

Section 3.2 provides a brief review of the literature. Section 3.3 layouts the model and the results. Section 3.5 concludes.

3.2 Literature

Since D. W. Diamond and Dybvig (1983), liquidity risk has been the focus of many studies but most of the literature uses the simplifying assumption that banks only hold cash as liquid securities. We take a novel approach by extending the set of liquid securities available to the bank and by assuming that they can invest in liquid but risky securities. Assuming that liquid securities are perfectly safe is sufficient for the large part of the banking literature on liquidity that focuses on the role of banks in liquidity provision and liquidity transformation. D. W. Diamond and Dybvig (1983) explain how banks can provide liquidity to households while investing in long-term illiquid projects and their model argues in favor of deposit insurance to prevent

costly bank runs. D. W. Diamond and R. G. Rajan (2001) go one step further in understanding why long-term illiquid assets may be coupled with fragile liabilities such as demand deposits. They show that the fragility of bank liabilities disciplines the banks and enhance the value of long-term illiquid assets. The fact that banks are prone to liquidity problems leads Kashyap, R. G. Rajan, and Jeremy C. Stein (2002) to show that banks can economize on costly liquidity buffers by holding assets and liabilities with imperfectly correlated liquidity risk. The liquidity creation role and run prone characteristic play an essential part in understanding the interactions between assets and liabilities of financial intermediaries but these theories do not offer much insight about the optimal portfolio allocation and risk-shifting³.

The absence of the modeling of risky securities is also due to a lack of rationale for banks to hold marketable securities. Hanson et al. (2015) note that today's commercial banks are mainly funded with safe deposits but invest in risky loans and risky securities. They provide a possible explanation of why commercial banks are holding long-term illiquid loans and risky securities. A safe deposit structure is alike the liability focused view of banking started by Gorton and Pennacchi (1990) and arguing for a "narrow banking" system where the fundamental role of financial intermediaries is to create safe-like securities for depositors⁴. In Hanson et al. (2015), commercial banks and shadow banks are competing in the business of creating safe-like securities but they do so differently depending on their funding structures. Traditional banks are assumed to have a stable source of funding that gives them a comparative advantage at holding illiquid loans and risky securities. On the contrary, shadow banks are subject to runs and fire sales losses and are more likely to hold safe and short-term assets to back their liabilities. While the funding structure is most certainly an important determinant of the asset structure of financial intermediaries, it does not take into account two important functions that commercial banks perform, namely liquidity

³Repullo (2005) is a counter example with cash and endogenous risk-shifting on the long term asset. Recently, Calomiris, Heider, and Hoerova (2015) analyze risk management for a bank with cash holdings.

⁴See also Pennacchi (2012)

management and risk management.

While simplistic in essence, viewing banks as holding risky illiquid and liquid assets raises basic yet fundamental questions. The one that is the focus of this paper is to understand the risk behavior on the overall bank's balance sheet, that is the aggregate portfolio choice of illiquid and liquid securities, and more specifically the aggregate correlation between illiquid and liquid assets. This question is very much like the traditional considerations of portfolio management. Indeed, considering banks as portfolio managers dates back to Pyle (1971) and O. D. Hart and Jaffee (1974). The portfolio approach explicitly considers the risk management performed by financial intermediaries by building on the classical mean-variance portfolio approach of Markowitz (1952). O. D. Hart and Jaffee (1974) show the existence of a separation theorem when the portfolio approach is applied to banks. That is, the mix of assets chosen by financial intermediaries is independent of the parameters of the utility function and can be logically separated from the decision on the size of the portfolio. This approach is the natural tool to study the effects of capital requirement on risk shifting. This has been done, among others, by Koehn and Santomero (1980), Kim and Santomero (1988) and Rochet (1992). Kim and Santomero (1988) shows that an increase in capital requirements does not necessarily decrease a bank's probability of failure because of portfolio reshuffling. Kim and Santomero (1988) argues that risk-weighted capital requirements must be used if one wants such regulation to be effective, and Rochet (1992) shows that capital regulations are effective only when banks behave as portfolio managers as opposed to value maximizers, highlighting again the need for a risk-weighted approach.

One weakness of the portfolio approach is that it fails to capture assets heterogeneity in their liquidity dimension. As a result, studies focusing on liquidity risk as has been using a simpler approach, while abstracting from standard portfolio decisions, such as correlations⁵. In a sense this paper lies in between liquidity management and risk management by shedding lights on

⁵Acharya and Pedersen (2005) provide an asset pricing model considering a different type of liquidity risk: the risk of not being able to easily sell a security. I study here liquidity risk coming from liabilities. In the taxonomy of Brunnermeier and Pedersen (2009), I focus on funding liquidity as opposed to market liquidity.

portfolio choices and risk-shifting coming from liquidity risk. It extends the traditional models of liquidity risk in the vein of D. W. Diamond and Dybvig (1983) by relaxing the bank's set of liquid assets and introducing the choice of correlation between illiquid and liquid assets. By doing so, it analyses a type of risk referred to as *aggregate asset risk* that has not been formally studied before.

3.3 The Model

The model features an economy with three dates $t \in \{0, 1, 2\}$, one good, and two types of agents: banks and households.

Households There is a continuum of households of size one, each endowed with one unit of good at time 0 that can be consumed at time 1 or 2. As in D. W. Diamond and Dybvig (1983), each household faces the risk of being an early or a late consumer. There is aggregate uncertainty⁶ in the economy represented by a state of nature s that can take two values, H and L, with equal probabilities. The probability of being an early consumer in state s is denoted by λ_s where

$$0 < \lambda_L < \lambda_H \leq 1$$

Aggregate uncertainty is resolved at date 1 when the state s is publicly observed. We denote p_s the probability of state s to realize. The discount rate is normalized to 0 and households are risk neutral with utility

$$u_s(c_1, c_2) = \begin{cases} c_1 & \text{w.pr. } \lambda_s \\ c_2 & \text{w.pr. } (1 - \lambda_s) \end{cases}$$

Where c_t represents consumption at date t . There exists investment opportunities in the economy but we assume that households do not have the necessary skills to undertake them. Rather, financial intermediation is essential in that households invest their endowment in a bank that undertakes investment opportunities on their behalf.

⁶Aggregate uncertainty is as in Allen and Gale (2007).

Banks There is one bank that is assumed to be representative of the banking sector. The bank has no initial endowment and issues financial claims to households. It issues short term demand deposits that can be claimed by households at any time 1 or 2 as well as equity capital. We assume that deposits are ensured so that households do not require a risk-premium for holding risky deposits. Because the banking sector is the only intermediary in the economy, it absorbs all households initial endowments. The overall size of the banking sector is thus one. The bank's capital structure is exogenously fixed with k in capital and $(1 - k)$ in deposits. An exogenous capital structure is a strong assumption. It is true if any capital ratio imposed by a regulatory body is binding in equilibrium, which is assumed here. This assumption prevents us to understand the rationale of banking regulation that has to be taken as given. Among others, Allen and Gale (2007) and Rochet (1992) analyze this question. The model can be generalized by introducing short term debt along with deposits as liabilities. The results go through as long as there is uncertainty on the amount of liquidity needed at the interim date, time 1. In this model, the stochasticity of the liquidity shock comes from uncertain deposits withdrawals but it can arise from short-term debt combined with uncertain access to refinancing.

At time 0, the bank provides firm lending. Loans are risky and return R or 0 with equal probabilities. Without loss of generality, we assume that all loans are perfectly correlated ; they either all return R or all return 0. Loans are illiquid in the sense that the bank cannot sell them on the market without incurring fire sales losses that will be specified later.

At time 1, aggregate uncertainty about household preferences is resolved and a fraction of depositors withdraw their deposits. It is assumed that external financing is infinitely costly for banks at that date. Therefore, the bank can only rely on existing assets to face depositors withdrawals. If the bank only holds firms loans, it is subject to fire sales losses. To prevent this adverse effect, the bank can invest in assets that provide liquidity at the interim date. We denote them *securities*. The focus of the paper is to study an economy in which perfectly safe assets are not available. It means that the bank cannot hold assets that are providing liquidity in every

future state of the world. Therefore we assume that securities provide either r or 0 units of liquidity at the interim date with equal probabilities. This assumption can be interpreted in two ways. First, one may think of securities as being easily marketable assets that provide uncertain returns, such as risky securitized assets. Even if these assets can be sold at their fundamental value, their ability to generate liquidity is state dependent because of their specific risk. Second, the ability for an asset to provide liquidity does not only depend on the asset alone, but also on the macroeconomic environment or the willingness of economic agents to trade. For instance, the value of government bonds that can be considered safe depends on monetary policy and interest rates. Also, market sentiment can quickly turn a liquid security into a very illiquid asset. One can think of episodes of liquidity dry ups during the 2007-2008 financial crisis. The key point is that it is very difficult for economic agents such as banks to know with certainty the future ability of assets to provide liquidity, which by construction makes liquidity uncertain. Here, we assume that future liquidity is stochastic but that its distribution is known ; there is no ambiguity about the liquidity that securities provide in future periods. Also, we assume that the bank and market participants behave rationally. Assuming that perfectly safe assets are non-existent implicitly assumes markets incompleteness. However, even if markets are complete, it can be shown that banks optimally use risky securities as long as they are sufficiently cheaper than safe assets. This condition is likely to be true in today's economic environment in which safest assets are in the negative rates territory.

At time 0, the bank chooses the amount to invest in loans and liquid assets. It lends $(1 - \gamma)$ to firms and buys γ securities. By construction, the bank cannot choose the individual risk of securities and loans. Instead, it faces a panel of borrowing firms and needs to choose how much to lend to each firm. This flexibility allows the bank to adjust the correlation between the returns of the loans portfolio and the liquidity provided by the securities. Assume the bank holds a portfolio of securities backed by real estate assets and that it can lend to either a construction or an agricultural firm. The risk on the real estate markets creates uncertainty on future liquidity provided by

the securities. It is likely that the creditworthiness of the construction firm is positively correlated with securities liquidity. However the agricultural firm is not impacted by real estate market uncertainty. By choosing between lending to the construction or to the agricultural firm the bank can influence how its access to liquidity covaries with future expected loan returns. Instead of modeling the underlying portfolio choice with a set of securities and firms, we express the problem in a reduced form, and tie securities and loans with an endogenously chosen correlation. The probability of loans to return R conditional on securities providing r units of liquidity is given by:

$$P[\tilde{R} = R | \tilde{r} = r] = \frac{1}{2}(1 + \rho)$$

Where ρ is a proxy of the correlation between loan returns and securities liquidity. For simplicity we focus on positive correlations and assume

$$0 < \rho < 1$$

For the problem to be interesting, there needs to be a cost of holding securities. We assume that investing in long term real assets is more profitable than holding liquidities:

$$R > r$$

At time 1, the bank is subject to a liquidity shock on its deposits. It faces withdrawals of a fraction λ_H of its deposits with probability p_H , and λ_L with probability $p_L = 1 - p_H$. Because securities can be sold at their fundamental values the bank uses them in priority to face the liquidity shock. If not enough liquidity is available through the sale of securities, it is forced to sell long term loans on the market. Securities liquidity and loan returns are covarying so the value of the loans are impacted by the state of the economy as well as the ex-ante choice of correlation. If securities end up providing low liquidity and loan returns are highly correlated with liquidities, their value is low and the bank needs to sell a large amount of them to face the liquidity shock. The bank defaults if it cannot raise enough liquidity by selling securities and loans combined. Otherwise it carries the remaining loans to time 2.

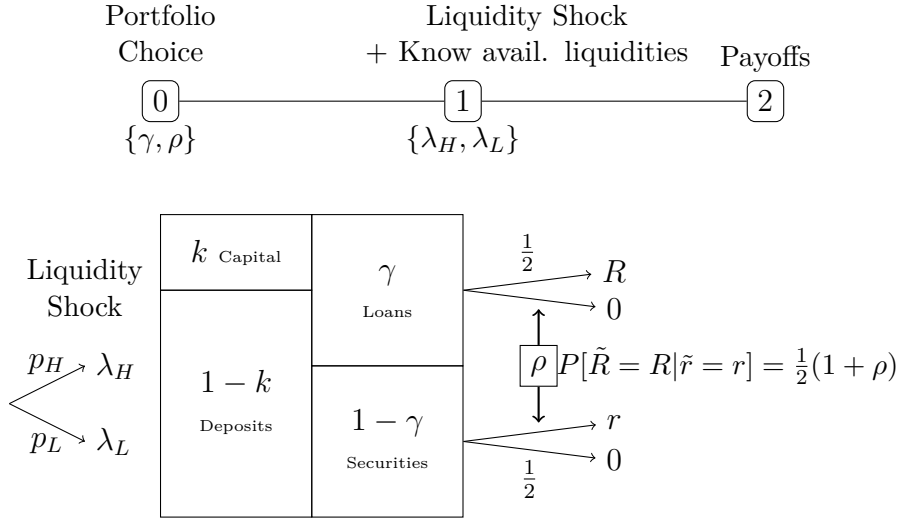


Figure 3.1: Setup

The bank maximizes shareholders value, that is time 2 equity value. Figure 3.1 represents the timeline of the model as well as the bank’s balance sheet structure.

Fire sales We assume that there is a price impact of selling loans on the market. The more loans the bank needs to sell, the lower their unit price. It captures the endogeneity of fire sales prices as in the models of Shleifer and Vishny (1992), Allen and Gale (1994), D. W. Diamond and R. G. Rajan (2011) or Jeremy C Stein (2012). Fire sales are such that if the bank wants to raise x units of liquidity, it needs to sell $F(x)$ worth of time 2 expected loan returns. We make the following assumption regarding the fire sales.

Assumption 3.

$$F(x), F'(x), F''(x) > 0 \tag{3.1}$$

$$F(0) = 0 \tag{3.2}$$

$$F'(0) > r \tag{3.3}$$

Assumption 3 states that F is positive, increasing and strictly convex. Moreover, it forces fire sales to be costly enough so that it is always sub-

optimal to use illiquid assets instead of liquid securities as liquidity provider.

A linear F is akin an exogenous fire sales discount independent of the number of sold illiquid assets. Price impact is captured by forcing the strict convexity of F and price impact losses are captured by the following expression:

$$L(x) = xF'(x) - F(x)$$

$L(x)$ represents the loss due solely to price impact. A linear F would result in $L(x) = 0$. Note that the price impact loss is strictly increasing in the amount of loans sold as $L'(x) = F'(x) > 0$.

Intuitions Before diving into the resolution, it seems important to describe the general mechanisms behind the results. The focus of the paper is to understand the endogenous choice of aggregate asset risk as defined above, that is, the correlation between the portfolios of illiquid loans and liquid securities. To that end, we try to capture important determinants of this choice of risk and denote them liquidity management and risk management. Liquidity management aims at supporting long term assets⁷, or loans, as well as avoiding liquidity shortages. Because the bank is subject to limited liability, liquidity management gives incentives for the bank to correlate liquidity with loan returns and thus increase aggregate asset risk. Indeed, it is profitable for the bank to support long term assets when these assets are the most valuable. It incentivizes the bank to secure more liquidity in the states in which loans have high returns. The effect is much like standard risk-shifting except that it arises from liquidity considerations and not purely from capital structure. In fact we will see that unlike traditional risk-shifting, aggregate asset risk decreases with leverage. Risk management is usually aiming at reducing ex-post variance in returns, as in Froot, David S Scharfstein, and Jeremy C. Stein (1993) or Froot and Jeremy C. Stein (1998). In our context, the goal of risk management is to mitigate the cost of liquidity shortages that are embedded in fire sales costs. While captured differently, the effects are identical as reducing fire sales losses is attained by lowering the variance be-

⁷As in Holmström and Tirole (2011)

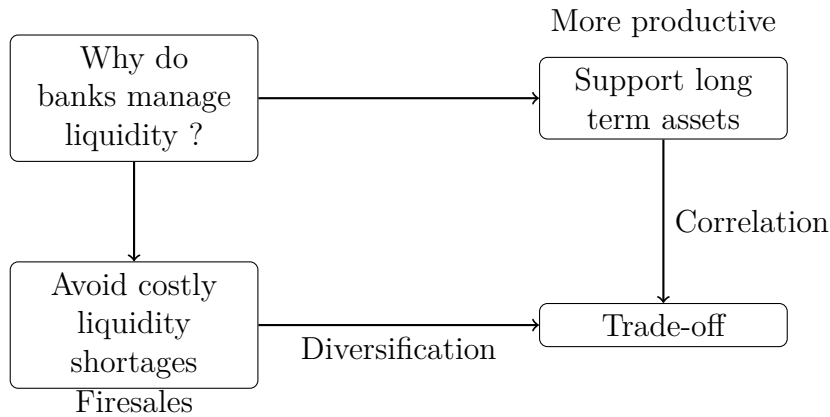


Figure 3.2: Trade-off

tween liquidity provisions and loan returns. To decrease fire sales losses, the bank tries to enhance loans value when liquidity is scarce. This is achieved by decreasing the correlation between loan returns and liquidity provision. Risk management in this model incentivizes the bank to decrease aggregate asset risk. It is the trade-off between these two forces that gives rise to an endogenous choice of aggregate asset risk. Both channels are illustrated in figure 3.2.

Now that we described all agents and the fire sales mechanism, we can analyze the bank choices. We solve the model by backward induction. All proofs are provided in the appendix.

3.3.1 Liquidity shock

At time 1, the bank holds a fraction γ of securities and needs to pay $\lambda_s(1-k)$ to early depositors, where s relates to a high (λ_H) or low (λ_L) liquidity shock. Securities are natural providers of liquidity and are first used to pay depositors. If not enough liquidity is available through securities, the bank sells loans on the market and is subject to a price impact loss from fire sales.

When securities provide r units of liquidity, the bank can repay up to γr to depositors at no cost. More early depositors means the bank needs to raise an additional $\lambda_s(1-k) - \gamma r$ units of liquidity by selling illiquid loans. It can either have enough loans and pay back all depositors at time 1 or default. If

liquidity is scarce ($\tilde{r} = 0$), the bank has to raise the full $\lambda_s(1 - k)$ through loans sales. Lemma 7 defines the thresholds of early depositors at which the bank is subject to fire sales or defaults.

Lemma 7. *There exists $\lambda_0 < \lambda_2$ and $\lambda_1 < \lambda_2$, such that, for a given liquidity shock state s ,*

- *When securities provide r units of liquidity:*
 - *If $\lambda_s < \lambda_0$, the bank has enough liquidity to face withdrawals from depositors and there are no fire sales.*
 - *If $\lambda_0 < \lambda_s < \lambda_2$, the bank needs to sell an amount $F(M_r^s)$ of loans.*
 - *If $\lambda_s > \lambda_2$, the bank defaults.*

$$\text{Where } M_r^s = \frac{\lambda_s(1-k) - \gamma r}{\frac{1}{2}(1+\rho)R}$$

- *When securities provide no liquidity:*
 - *If $\lambda_s < \lambda_1$, the bank needs to sell an amount $F(M_0^s)$ of loans.*
 - *If $\lambda_s > \lambda_1$, the bank defaults.*

$$\text{Where } M_0^s = \frac{\lambda_s(1-k)}{\frac{1}{2}(1-\rho)R}$$

When securities provide r units of liquidity, the expected value of one unit of loan is $\frac{1}{2}(1+\rho)R$. The bank needs to sell an amount $\max\left\{0, F\left(\frac{\lambda_s(1-k) - \gamma r}{\frac{1}{2}(1+\rho)R}\right)\right\}$ of loans to obtain extra liquidity. The bank defaults if $(1-\gamma) < F\left(\frac{\lambda_s(1-k) - \gamma r}{\frac{1}{2}(1+\rho)R}\right)$, defining the threshold λ_2 . If securities provide no liquidity, the expected value of one unit of loan is $\frac{1}{2}(1-\rho)R$ and the bank has to sell $F\left(\frac{\lambda_s(1-k)}{\frac{1}{2}(1-\rho)R}\right)$ units of loans. The bank defaults if $(1-\gamma) < F\left(\frac{\lambda_s(1-k)}{\frac{1}{2}(1-\rho)R}\right)$, defining the threshold λ_1 . It is trivial to see that $\lambda_1 < \lambda_2$ as long as ρ is large enough, and it is always true when $\rho > 0$.

We assume that the bank does not always default when liquidity is scarce. That is, the low liquidity shock is low enough so that the bank can survive by selling illiquid loans.

Assumption 4.

$$\lambda_L < \lambda_1$$

We can now turn to time 0 portfolio choices.

3.3.2 Portfolio choice

At time 0, the bank needs to choose both the investment mix between liquid and illiquid assets, γ , and the correlation between the two, ρ , also referred to as *aggregate asset risk*.

Before laying out the bank's objective function, we can restrict the set of acceptable choices for γ . Holding liquidity is costly because illiquid assets are more productive. It is thus sub-optimal for the bank to hold more liquid assets than what is necessary to face the highest liquidity shock λ_H . Equivalently, liquid securities allow the bank to carry illiquid assets to maturity. Using illiquid loans as providers of liquidity is always more costly than using securities, as is ensured by assumption 3. Therefore it is sub-optimal to use illiquid loans as liquidity buffer at time 1, and the bank must at least hold enough liquid assets to face the lowest liquidity shock λ_L .

Lemma 8 formalizes these results.

Lemma 8. *The optimal mix of assets γ^* is such that*

$$\underline{\gamma} \equiv \frac{\lambda_L(1-k)}{r} \leq \gamma^* \leq \frac{\lambda_H(1-k)}{r} \equiv \bar{\gamma}$$

which is equivalently expressed as

$$\lambda_L \leq \lambda_0 \leq \lambda_H$$

It is important to note that the thresholds $\lambda_0, \lambda_1, \lambda_2$ defined in lemma 7 are dependent upon the bank choices. In fact, it is reasonable to expect that the ability of the bank to impact its default probabilities is an important determining factor of the choice of γ and ρ . We now show that the bank *endogenously* defaults in the worst state of the world. At the optimum, the bank choices are such that it always defaults in the high shock state if

liquidity is scarce, and it never defaults if securities provide liquidity. This is formalized in lemma 9.

Lemma 9. *The optimum bank's portfolio is such that*

$$\lambda_1 < \lambda_H < \lambda_2$$

This result is central for the mechanisms at play as it creates a non null *endogenous* probability that the bank defaults if securities do not provide liquidity. Combined with limited liability, it creates risk-shifting incentives.

We can now express the bank's objective function. The expected value of time 2 bank's equity, denoted V , is given by

$$\begin{aligned} V = & p_L \left\{ \gamma r + (1 - \gamma) \frac{1}{2} (1 + \rho) R - \lambda_L (1 - k) + [(1 - \gamma) - F(M_0^L)] \frac{1}{2} (1 - \rho) R \right\} \\ & + p_H \left\{ [(1 - \gamma) - F(M_r^H)] \frac{1}{2} (1 + \rho) R \right\} \\ & - p_L (1 - \lambda_L) (1 - k) - p_H (1 - \lambda_H) (1 - k) \end{aligned} \tag{3.4}$$

Where M_0^L and M_r^H are the amounts of loans that need to be sold in states where securities do not provide liquidity and when the liquidity shock is low (M_0^L), or when securities provide liquidity and when the liquidity shock is high (M_r^H).

The first line corresponds to the expected equity value in the low liquidity shock state. When securities provide liquidity (3 first terms), there are no fire sales, the bank has enough liquidity to pay depositors, and all illiquid assets ($(1 - \gamma)$) are carried up to time 2. When securities do not provide liquidity (last term), the bank needs to raise M_0^L by selling illiquid loans. The second line is the equity value in the high liquidity shock state. By lemma 9, the bank defaults if liquidity is scarce. Otherwise it needs to sell M_r^H units of loans to pay early depositors. The last line is simply the payment of late consumers at time 2. Figure 3.3 illustrates the different states of the world at time 1.

Note that if the bank does not default at time 1, it is assumed to pay back late depositors in full. This is not realistic if loan returns are insufficient in

	Low shock (λ_L, p_L)	High shock (λ_H, p_H)
High Liq. ($\tilde{r} = r, \frac{1}{2}$)	Enough Liquidity No Firesales	Firesales Sell for M_r^H in loans
Low Liq. ($\tilde{r} = 0, \frac{1}{2}$)	Firesales Sell for M_0^L in loans	Default

Figure 3.3: States of the world at time 1

the last period. Apart from simplicity, this assumption ensures that limited liability in the last period does not affect the results. We are interested in the effects of the liquidity shock alone, and limited liability at time 1 is at the root of risk-shifting. However, we do not aim at capturing the effects of limited liability in the last period. Traditionally, leverage affects risk-shifting on the loans portfolio if there is limited liability at time 2. This effect is of no interest here. In fact the model does not allow to capture risk-shifting on the loans portfolio because risk is fixed on individual portfolios by assumption. Hence, removing the limited liability assumption at time 2 does not affect the generality of the results.

Risk-Taking trade-off The value function clearly shows the risk-taking trade-off. It can be broken down into risk-loving incentives in the high shock state and diversification incentives in the low shock state.

In the high liquidity shock state, the bank does not default only if securities provide liquidity ($\tilde{r} = r$). Limited liability in that state creates risk-loving incentives, and the bank can increase time 2 equity value by holding illiquid assets whose returns are correlated with its securities. Three effects are at play: increasing correlation allows the bank to sell fewer loans, reduces the price impact loss, and increases loan expected returns. A similar mechanism appears in the low liquidity shock state where the bank incurs fire sales only when securities do not provide liquidity. Decreasing the correlation reduces the fire sales losses by decreasing the amount of loans to sell, reducing the

price impact loss, and increasing expected returns. This trade-off only appears because fire sales can occur in both the high and the low shock state. Therefore, normalizing λ_L to zero would remove the diversification incentives and result in a correlation always at 1⁸.

The bank optimizes expected equity value by choosing the mix of assets as well as the correlation between the portfolios of loans and securities:

$$\max_{\gamma, \rho} V(\gamma, \rho) \quad (3.5)$$

$$\text{s.t. } \underline{\gamma} < \gamma < \bar{\gamma} \quad (3.6)$$

Where V is as in equation 3.4 and inequation 3.6 comes lemma 8. Lemma 10 characterizes the solution.

Lemma 10. *The optimal choices of mix of assets γ^* and correlation ρ^* follow the following equations:*

$$\gamma^* = \max \left[\underline{\gamma}, \bar{\gamma} - \frac{1}{2r}(1 + \rho^*)RF'^{-1} \left(\frac{1}{r} \left[\frac{1}{2}(1 + \rho^*)R + (R - r)\frac{p_L}{p_H} \right] \right) \right] \quad (3.7)$$

$$p_H(1 - \gamma^*) + [p_H L(M_r^H(\gamma^*, \rho^*)) - p_L L(M_0^L(\rho^*))] = 0 \quad (3.8)$$

Where $L(x) = xF'(x) - F(x)$

Lemma 10 reveals that the risk-taking trade-off described previously creates a trade-off between the mix of assets and correlation that is at the basis of the subsequent results on regulatory implications. Equation 3.7 is the first order condition for γ . It shows that optimal security holdings are inversely related to aggregate asset risk. That is, the higher the correlation between the portfolios of loans and securities, the lower the amount of securities held by the bank. The first order condition for ρ (equation 3.8) shows an identical relationship. To see it, note that M_r^H is decreasing in both ρ and γ and M_0^L is increasing in ρ . Therefore, any decrease in γ^* has to be compensated by an increase in ρ^* . The intuitions are the following. On the one hand, when aggregate asset risk is higher, gains for holding loans are increasing because

⁸This is because diversification incentives are only captured by fire sales. One could think of a model where diversification comes from a time 2 concave investment opportunity.

it increases expected returns when securities provide liquidity. Expected returns decrease when liquidity is scarce in the low shock state, but can be compensated by holding more loans. On the other hand, when the bank holds more securities, it reduces fire sales losses in the high shock state but not in the low shock state. One way to compensate the loss in the low shock state is to increase the correlation thus lowering fire sales in that state.

Lemma 11. *aggregate asset risk and securities holdings are inversely related.*

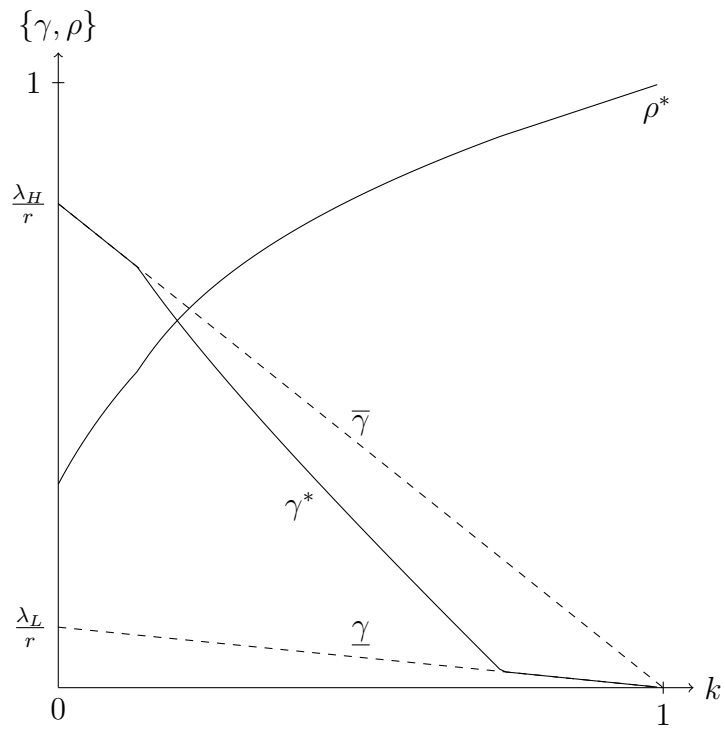
Let us illustrate the bank's optimal choices by numerically solving the model. We use the functional form $F(x) = e^{rx} - 1$ for the fire sales. Figure 3.4 plots the optimal mix of assets γ^* and correlation between loans and securities ρ^* as a function of the leverage. It clearly shows the trade-off between the mix of assets and aggregate asset risk.

We have seen that liquidity risk combined with fire sales create a risk-taking trade-off. Moreover, this trade-off leads to an inverse relationship between security holdings and aggregate asset risk taking. Next section draws regulatory implications for two of the most important tools used in banking regulation that are capital and liquidity requirements.

3.4 Regulatory Implications

This model is a partial equilibrium approach of optimal portfolio holdings. As it does not formally include a regulator with a social objective function, it stays silent on the rationale for regulation as well as on optimal regulation. Nevertheless, by shedding lights on how banks manage their assets when facing liquidity risk, it allows us to understand how banks respond to regulatory changes. We focus our analysis on capital and liquidity requirements.

Figure 3.4 paves the way for intuitions. It first shows that a bank with a higher capital ratio has a higher equilibrium aggregate asset risk. It also illustrates the inverse relationship between securities holding and aggregate asset risk which says that forcing banks to hold more liquid assets decreases aggregate asset risk. Finally, it suggests that the sensitivity of the effects of liquidity requirements on aggregate asset risk is higher for more levered



The fire sales function is $F(x) = e^{rx} - 1$ and the parameters are $p_L = .8, p_H = .2, R = 1.05, r = 1, \lambda_L = .1, \lambda_H = .8$.

Figure 3.4: Optimal mix of assets and correlation as a function of leverage.

banks. That is, ρ decreases more as a result of an increase of γ for low values of k . We analyze in turn these intuitions.

3.4.1 Capital Regulation

Historically, the main tool in micro and macro prudential regulations is the use of capital requirements, introduced in Basel I and Basel II regulatory frameworks. Lowering the leverage of financial intermediaries aims at reducing their default probability. Additionally, traditional risk-shifting mechanism suggests that more levered institutions have an acute appetite for risk, which may not be socially optimal. This is true in traditional corporate finance models that do not take into account liquidity risk. In banking, most models focus on the risk of illiquid asset returns, loans in my model. Here we study a specific type of risk that captures how liquidity provisions are covarying with asset returns, and we find that this risk – aggregate asset risk – increases with capital ratio.

Proposition 5. *An increase in capital requirements increases aggregate asset risk.*

Increased capital requirements lowers liquidity risk by decreasing the size of the liquidity shocks. This creates two channels leading to an increase in aggregate asset risk. The first is a decrease in fire sales costs in both the high and the low liquidity shock states. Fire sales are therefore less costly and the need to diversify the portfolio becomes less important in the low shock state. Risk-shifting incentives, however are still present in the high shock state due to limited liability. The second effect is an increase in loan investment. Because shocks are smaller, the need for liquid assets decreases and the bank invests more in long-term assets. However, an increase in loans increases the benefits of correlation. Because the bank holds less liquidity in the high liquidity shock state, it has to sell more illiquid assets, and a way to mitigate this effect is for these assets to be worth more. This can be done by increasing the correlation. In the low shock state, the mix of assets is irrelevant because fire sales only occur when securities do not provide liquidity, hence it is as if the bank does not hold any.

Another way to interpret this result is to keep in mind the inverse relationship between securities holding and aggregate asset risk-taking. A capital increase lowers liquidity shocks and decreases securities holdings. As a result, it increases aggregate asset risk. Figure 3.4 illustrates this result.

3.4.2 Liquidity Regulations

Liquidity Coverage Ratio (LCR) is a recent tool in banking regulation that has been introduced as part of the Basel III framework. It aims at controlling the amount of liquid assets relative to deposits. In the context of the model, it means that the regulator can fix the ratio $\xi = \frac{\gamma}{1-k}$.

Introducing liquidity requirements simplifies the model, as it renders the mix of assets exogenous. The bank's choice of aggregate asset risk is thus only governed by equation 3.8, the first order condition for ρ . We find that liquidity requirements have a positive effect on aggregate asset risk taking.

Proposition 6. *An increase in the Liquidity Coverage Ratio decreases aggregate asset risk taking.*

We find that an increase in the LCR has a positive impact on aggregate asset risk. Asking the bank to hold more securities has a beneficial impact on fire sales in the high shock state. Because the bank does not need to rely as much on assets sales, it becomes costly to hold correlated assets if the low shock state realizes relative to the benefits in the high shock state. That gives incentives for decreasing the overall correlation. In addition, by limiting investment in illiquid assets, the bank reduces the profits of surviving the high shock state, which tilts even more the bank's choice towards uncorrelated assets.

Both capital and liquidity requirements aim at reducing the bank's default probability. However they do so in fundamentally different ways that impact aggregate risk-shifting incentives. Capital requirements affect all future states of the world and reduce all future liquidity shocks equally. It turns out that it has an adverse effect on aggregate asset risk. However, liquidity requirements are lowering liquidity shocks only in states where they are par-

ticularly severe. As a result, it is a much more effective tool for mitigating aggregate asset risk.

Figure 3.4 also suggests that the effect of liquidity requirements are altered by leverage. That is, the impact of liquidity requirements are different depending on the leverage. This is a very important implication for regulators. It means that what matters is the *joint* choice of capital and liquidity requirements, and this model highlights the interactions between both regulatory tools.

Proposition 7. *The impact of liquidity requirements on aggregate asset risk is higher for more levered banks.*

When leverage is high (small k), liquidity risk is at its highest as shocks are large. Consequently, fire sales costs are more sensitive to changes in expected loans sales, leading to a greater impact of liquidity requirements on aggregate risk shifting.

3.4.3 Impact of liquidity shock

Finally, we try to explore how the characteristics of the liquidity shock might impact bank's liquidity management.

The model features aggregate uncertainty in the economy and one representative bank. That construction implicitly assumes that idiosyncratic risk is netted out and that the bank only faces systemic risk. However, the banking system is not as simple, and there is heterogeneity in the type of shocks faced by individual banks. For instance, it is likely that small banks face higher idiosyncratic risk than bigger banks. It might therefore be important to understand what the model has to say on any cross-sectional implications of liquidity management and aggregate risk taking.

I do so here in a very simple and reduced form way, by assuming that idiosyncratic risk influences the distribution of the liquidity shock.⁹ More specifically, we assume that idiosyncratic risk increases the volatility of the

⁹It is also quite intuitive to expect that fire sales differ substantially when a systemic or idiosyncratic shock hits. I do not try to capture such effects here as it would require a much deeper analysis than what the current model can provide.

liquidity shock. We find that small banks are more likely to hold more liquid assets but also to take on higher level of aggregate asset risk.

Proposition 8. *Banks that are facing more idiosyncratic risk hold more liquid assets but have higher aggregate asset risk profiles.*

When the high shock state becomes more prevalent, fire sales losses are more likely to occur and holding liquidity is beneficial. However, if the large liquidity shock is more probable, limited liability increases risk shifting incentives, leading to a higher choice of correlation.

This result is interesting in the light of Basel 3 applicability of the LCR requirements in the United States. Liquidity requirements in the US only apply to banks whose total assets are larger than \$250 billions, mainly large banks. However, a quick look at the FIDC insured bank's assets as of 2012 reveals that more than 40% of total banking assets are held by banks not subject to liquidity requirements. Proposition 8 highlights the need for increasing regulatory focus on small banks that have traditionally been ignored, as regulators around the world have been more concerned about systemic risk.

I have shown that capital ratio regulations tend to increase aggregate risk-taking while liquidity regulations tend to decrease it. In addition, the effectiveness of liquidity requirements is higher when banks are more levered. We have also highlighted the fact that small banks may be more affected by the mechanisms revealed in this paper, suggesting that imposing only capital regulations on small banks may not be enough if one is concerned about aggregate risk shifting.

3.5 Conclusion

In this paper, we propose a banking model of optimal portfolio choice combining illiquid assets and liquid securities. We depart from traditional banking models by assuming that bank securities provide stochastic future liquidity. Introducing uncertainty in future liquidity raises the question of how banks manage the risk between liquidity provisions and long term asset returns. This risk is defined as being the correlation between illiquid asset returns

and liquidity provisions from securities, and is denoted *aggregate asset risk*. The presence of liquidity risk in the model generates risk-loving incentives that are balanced by diversification motives created by fire sales. The model reveals that this trade-off gives rise to an inverse relationship between security holdings and aggregate risk taking. In other words, the more liquidities the bank holds, the less correlated its securities and long-term loans are. However, the bank tends to correlate its assets when it has more capital. It naturally leads me to show that current banking regulatory tools such as capital requirements and liquidity requirements have opposite effect on bank's appetite for aggregate asset risk. While imposing high capital ratio leads to higher aggregate risk taking, liquidity requirements are effective in decreasing risk-taking. The reason is that capital ratio affects all future liquidity shocks in an identical manner while liquidity requirements have the ability to dampen liquidity shocks when they are the most severe. Finally, we draw the cross-sectional prediction that small banks are potentially more subject to take on aggregate asset risk due to their increased exposure to idiosyncratic risk. Overall, these results show that there is a tension between capital and liquidity requirements. That is, both have opposite effects on aggregate risk taking. It therefore calls for cautious regulatory design, and suggests that a regulatory framework that imposes high capital ratio and high liquidity requirement may not be optimal if one is concerned about aggregate asset risk.

Bibliography

- Acharya, Viral V. and Lasse Heje Pedersen (2005). “Asset Pricing with Liquidity Risk”. *Journal of Financial Economics* 77.2, pp. 375–410.
- Acharya, Viral V. and Philipp Schnabl (May 1, 2009). “How Banks Played the Leverage “Game””. *Financial Markets, Institutions & Instruments* 18.2, pp. 144–145.
- Allen, Franklin and Douglas Gale (1994). “Limited Stock Market Participation and Volatility of Asset Prices”. *American Economic Review* 84.4, pp. 933–955. JSTOR: 2118039.
- (2007). *Understanding Financial Crises*. Oxford University Press.
- Basel Committee (2013). “Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools”. *Basel Committee on Banking Supervision* (January), pp. 1–75.
- Brunnermeier, Markus K. and Lasse Heje Pedersen (2009). “Market Liquidity and Funding Liquidity”. *Review of Financial Studies* 22.6, pp. 2201–2238.
- Calomiris, Charles W., Florian Heider, and Marie Hoerova (2015). “A Theory of Bank Liquidity Requirement”.
- Cerasi, Vittoria and Sonja Daltung (Oct. 2000). “The Optimal Size of a Bank: Costs and Benefits of Diversification”. *European Economic Review* 44.9, pp. 1701–1726.
- Chemla, Gilles and Christopher A. Hennessy (Aug. 1, 2014). “Skin in the Game and Moral Hazard”. *The Journal of Finance* 69.4, pp. 1597–1641.
- Diamond, Douglas (1984). “Financial Intermediation and Delegated Monitoring”. *The Review of Economic Studies* 51.3, pp. 393–414. JSTOR: 2297430.

- Diamond, Douglas W. and Philip H. Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity”. *Journal of Political Economy* 91.3, p. 401.
- Diamond, Douglas W. and Raghuram G. Rajan (2001). “Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking”. *Journal of Political Economy* 109, pp. 287–327.
- (2011). “Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes”. *Quarterly Journal of Economics* 126.2, pp. 557–591.
- Freixas, Xavier and Jean-Charles Rochet (2008). *Microeconomics of Banking*. The MIT Press.
- Froot, Kenneth A., David S Scharfstein, and Jeremy C. Stein (1993). “Risk Management: Coordinating Corporate Investment and Financing Policies”. *The Journal of Finance* 48, pp. 1629–1658.
- Froot, Kenneth A. and Jeremy C. Stein (1998). “Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach”. *Journal of Financial Economics* 47.1, pp. 55–82.
- Gorton, Gary (2010). *Slapped by the Invisible Hand: The Panic of 2007*.
- Gorton, Gary and George Pennacchi (1990). “Financial Intermediaries and Liquidity Creation”. *The Journal of Finance* 45.1, pp. 49–71.
- (June 1, 1995). “Banks and Loan Sales Marketing Nonmarketable Assets”. *Journal of Monetary Economics* 35, pp. 389–411.
- Hanson, Samuel et al. (2015). “Banks as Patient Fixed Income Investors”. *Journal of Financial Economics*.
- Hart, Oliver D. and Dwight Jaffee (1974). “On the Application of Portfolio Theory to Depository Financial Intermediaries”. *The Review of Economic Studies* 41.1, pp. 129–147. JSTOR: 2296404.
- Hartman-Glaser, Barney, Tomasz Piskorski, and Alexei Tchisty (2012). “Optimal Securitization with Moral Hazard”. *Journal of Financial Economics* 104.1, pp. 186–202.
- Holmström, Bengt and Jean Tirole (2011). *Inside and Outside Liquidity*. MIT Press.
- Kashyap, Anil K., Raghuram G. Rajan, and Jeremy C. Stein (2002). “Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking”. *The Journal of Finance* 57.1, pp. 33–73.

- Kim, Daesik and Anthony M Santomero (1988). "Risk in Banking and Capital Regulation". *Journal of Finance* 43.5, pp. 1219–1233.
- Koehn, Michael and Anthony M Santomero (1980). "Regulation of Bank Capital and Portfolio Risk". *Journal of Finance* 35, pp. 1235–1244.
- Markowitz, Harry (1952). "Portfolio Selection". *The Journal of Finance* 7, pp. 77–91.
- Parlour, Christine A. and Guillaume Plantin (June 1, 2008). "Loan Sales and Relationship Banking". *The Journal of Finance* 63.3, pp. 1291–1314.
- Pennacchi, George (2012). "Narrow Banking". *Annual Review of Financial Economics* 4.0.
- Pyle, David H. (1971). "On the Theory of Financial Intermediation". *The Journal of Finance* 26.3, pp. 737–747.
- Repullo, Rafael (2005). "Liquidity , Risk Taking , and the Lender of Last Resort". *International Journal of Central Banking* 1.2, pp. 47–80.
- Rochet, Jean-Charles (1992). "Capital Requirements and the Behaviour of Commercial Banks". *European Economic Review* 36, pp. 1137–1170.
- Shleifer, Andrei and Robert W. Vishny (1992). "Liquidation Values and Debt Capacity: A Market Equilibrium Approach". *Journal of Finance* 47.4, pp. 1343–1366.
- Stein, Jeremy C (2012). "Monetary Policy as Financial Stability Regulation". *The Quarterly Journal of Economics* 127, pp. 57–95.

Appendix

Proofs

Lemma 7

Here, we compute the thresholds λ_0 , λ_1 and λ_2 of lemma 7. λ_0 is such that

$$\lambda_0 = \frac{\gamma r}{1 - k}$$

λ_1 is such that

$$(1 - \gamma) = F(M_0(\lambda_1)) \tag{3.9}$$

$$(1 - \gamma) = F\left(\frac{\lambda_1(1 - k)}{\frac{1}{2}(1 - \rho)R}\right) \tag{3.10}$$

$$F^{-1}(1 - \gamma) = \frac{\lambda_1(1 - k)}{\frac{1}{2}(1 - \rho)R} \tag{3.11}$$

$$\frac{1}{2}(1 - \rho)RF^{-1}(1 - \gamma) = \lambda_1(1 - k) \tag{3.12}$$

$$\lambda_1 = \frac{\frac{1}{2}(1 - \rho)RF^{-1}(1 - \gamma)}{1 - k} \tag{3.13}$$

λ_2 is such that

$$(1 - \gamma) = F(M_r(\lambda_2)) \quad (3.14)$$

$$(1 - \gamma) = F\left(\frac{\lambda_2(1 - k) - \gamma r}{\frac{1}{2}(1 + \rho)R}\right) \quad (3.15)$$

$$\frac{1}{2}(1 + \rho)RF^{-1}(1 - \gamma) = \lambda_2(1 - k) - \gamma r \quad (3.16)$$

$$\frac{1}{2}(1 + \rho)RF^{-1}(1 - \gamma) + \gamma r = \lambda_2(1 - k) \quad (3.17)$$

$$\lambda_2 = \frac{\frac{1}{2}(1 + \rho)RF^{-1}(1 - \gamma) + \gamma r}{1 - k} \quad (3.18)$$

$\lambda_0 < \lambda_2$ is by construction and it is trivial to see that $\lambda_1 < \lambda_2$.

Lemma 10

The bank optimizes the following program:

$$\begin{aligned} \max_{\gamma, \rho} p_L & \left\{ \gamma r + (1 - \gamma) \frac{1}{2}(1 + \rho)R - \lambda_L(1 - k) + [(1 - \gamma) - F(M_0^L)] \frac{1}{2}(1 - \rho)R \right\} \\ & + p_H \left\{ [(1 - \gamma) - F(M_r^H)] \frac{1}{2}(1 + \rho)R \right\} \\ & - p_L(1 - \lambda_L)(1 - k) - p_H(1 - \lambda_H)(1 - k) \\ \text{s.t. } & \underline{\gamma} < \gamma < \bar{\gamma} \end{aligned}$$

Assuming that we have an interior solution, γ^* follows the FOC

$$\begin{aligned} p_L(R - r) + p_H \left[\frac{1}{2}(1 + \rho)R \left(1 + \frac{\partial M_r^H}{\partial \gamma} F'(M_r^H) \right) \right] &= 0 \\ p_L(R - r) + p_H \left[\frac{1}{2}(1 + \rho)R \left(1 - \frac{r}{\frac{1}{2}(1 + \rho)R} F'(M_r^H) \right) \right] &= 0 \\ p_L(R - r) + p_H \left[\frac{1}{2}(1 + \rho)R - r F'(M_r^H) \right] &= 0 \end{aligned}$$

Solving for γ gives

$$\gamma^* = \bar{\gamma} - \frac{1}{2}(1 + \rho^*)RF'^{-1} \left(\frac{1}{r} \left[\frac{1}{2}(1 + \rho^*)R + (R - r) \frac{p_L}{p_H} \right] \right)$$

Which proves that $\gamma^* \leq \bar{\gamma}$.

ρ^* respects the FOC

$$p_L \left\{ \frac{1}{2}R(1 - \gamma) - \left[\frac{1}{2}R[(1 - \gamma) - F(M_0^L)] + \frac{\partial M_0^L}{\partial \rho} F'(M_0^L) \frac{1}{2}(1 - \rho)R \right] \right\} \\ + p_H \left\{ \frac{1}{2}R[(1 - \gamma) - F(M_r^H)] - \frac{\partial M_r^H}{\partial \rho} F'(M_r^H) \frac{1}{2}(1 + \rho)R \right\} = 0 \\ p_L \left\{ F(M_0^L) - \frac{\partial M_0^L}{\partial \rho} F'(M_0^L)(1 - \rho) \right\} + p_H \left\{ (1 - \gamma) - F(M_r^H) - \frac{\partial M_r^H}{\partial \rho} F'(M_r^H)(1 + \rho) \right\} = 0$$

We have

$$\frac{\partial M_0^L}{\partial \rho}(1 - \rho) = M_0^L \quad (3.19)$$

$$\frac{\partial M_r^H}{\partial \rho}(1 + \rho) = M_r^H \quad (3.20)$$

Plugging it into the FOC gives

$$p_H(1 - \gamma) + p_H[M_r^H F'(M_r^H) - F(M_r^H)] - p_L[M_0^L F'(M_0^L) - F(M_0^L)] = 0$$

Expressing the price impact loss as $L(x) = xF'(x) - F(x)$, we have

$$p_H(1 - \gamma) + p_H L(M_r^H) - p_L L(M_0^L) = 0$$

Lemma 11

Here we show that $\frac{\partial \gamma^*}{\partial \rho^*} < 0$ and that $\frac{\partial \rho^*}{\partial \gamma^*} < 0$. γ^* and ρ^* are jointly governed by equations 3.7 and 3.8. We define

$$g(\gamma^*, \rho^*) = \max \left[\underline{\gamma}, \bar{\gamma} - \frac{1}{2r}(1 + \rho^*)RF'^{-1} \left(\frac{1}{r} \left[\frac{1}{2}(1 + \rho^*)R + (R - r) \frac{p_L}{p_H} \right] \right) \right] - \gamma^* = 0 \quad (3.21)$$

$$h(\gamma^*, \rho^*) = p_H(1 - \gamma^*) + p_H L(M_r^H) - p_L L(M_0^L) = 0 \quad (3.22)$$

We have that

$$\frac{\partial g}{\partial \gamma^*} < 0, \frac{\partial g}{\partial \rho^*} < 0 \quad (3.23)$$

$$\frac{\partial h}{\partial \gamma^*} < 0, \frac{\partial h}{\partial \rho^*} < 0 \quad (3.24)$$

Applying the implicit function, we obtain

$$\frac{\partial \gamma^*}{\partial \rho^*} = -\frac{\partial g / \partial \rho^*}{\partial g / \partial \gamma^*} < 0 \quad (3.25)$$

$$\frac{\partial \rho^*}{\partial \gamma^*} = -\frac{\partial h / \partial \gamma^*}{\partial h / \partial \rho^*} < 0 \quad (3.26)$$

$$(3.27)$$

Lemma 9

We first prove that $\lambda_H < \lambda_2$ by showing that when $\lambda_H > \lambda_2$, $\frac{\partial V}{\partial \rho} > 0$. If this is true, λ_2 has to be such that $\lambda_2 \geq \lambda_H$. If $\lambda_H > \lambda_2$, the bank always defaults in the high shock state and its value function is

$$V = p_L \left\{ \gamma r + (1 - \gamma) \frac{1}{2} (1 + \rho) R - \lambda_L (1 - k) + [(1 - \gamma) - F(M_0)] \frac{1}{2} (1 - \rho) R \right\}$$

So we have

$$\frac{\partial V}{\partial \rho} = \frac{\partial M_0}{\partial \rho} \frac{1}{2} R F'(M_0) > 0$$

We now prove that $\lambda_1 < \lambda_H$ similarly by showing that when $\lambda_H < \lambda_1$, $\frac{\partial V}{\partial \gamma} > 0$. If true, it has to be that $\lambda_1 \leq \lambda_H$. If $\lambda_H < \lambda_1$, the bank never defaults in the high shock state and its value function is

$$V = p_L \left\{ \gamma r + (1 - \gamma) \frac{1}{2} (1 + \rho) R - \lambda_L (1 - k) + [(1 - \gamma) - F(M_0^L)] \frac{1}{2} (1 - \rho) R \right\} \quad (3.28)$$

$$+ p_H \left\{ [(1 - \gamma) - F(M_r^H)] \frac{1}{2} (1 + \rho) R + [(1 - \gamma) - F(M_0^H)] \frac{1}{2} (1 - \rho) R \right\} \quad (3.29)$$

So we have

$$\frac{\partial V}{\partial \rho} = \frac{\partial M_0^L}{\partial \rho} \frac{1}{2} RF'(M_0^L) p_L - \frac{\partial M_r^H}{\partial \rho} \frac{1}{2} RF'(M_r^H) p_H + \frac{\partial M_0^H}{\partial \rho} \frac{1}{2} RF'(M_0^H) p_H > 0$$

Note that these results are independent of the simultaneous choice of γ and ρ and also hold when γ is exogenous.

Proposition 5

To show that aggregate asset risk increase with capital requirements, we need to show that $\frac{d\rho^*}{dk} > 0$. ρ^* follows

$$p_H(1 - \gamma^*(k)) + p_H L(M_r^H(\gamma^*(k), \rho^*)) - p_L L(M_0^L(\gamma^*(k))) = h(k, \rho^*) = 0$$

We have

$$\frac{\partial \gamma^*}{\partial k} < 0, \frac{\partial M_r^H}{\partial \gamma^*} < 0, \frac{\partial M_0^* L}{\partial k} < 0$$

It follows that

$$\frac{\partial h}{\partial \rho^*} < 0 \tag{3.30}$$

$$\frac{\partial h}{\partial k} = -\frac{\partial \gamma^*}{\partial k} + p_L L'(M_r^H) \frac{\partial M_r^H}{\partial \gamma^*} \frac{\partial \gamma^*}{\partial k} - p_L L'(M_0^L) \frac{\partial M_0^* L}{\partial k} > 0 \tag{3.31}$$

Therefore

$$\frac{d\rho^*}{dk} = -\frac{\partial h / \partial k}{\partial h / \partial \rho^*} > 0$$

Proposition 6

We express the FOC for ρ with the replacement $\xi = \frac{\gamma}{1-k}$. ρ^* follows

$$h(\xi, \rho) = p_H(1 - \xi(1 - k)) + p_H L(M_r^H(\xi, \rho^*)) - p_L L(M_0^L(\rho^*)) = 0$$

We have

$$\frac{\partial h}{\partial \xi} < 0 \quad (3.32)$$

$$\frac{\partial h}{\partial \rho^*} < 0 \quad (3.33)$$

Applying the implicit function theorem gives

$$\frac{\partial \rho^*}{\partial \xi} = -\frac{\partial h / \partial \xi}{\partial h / \partial \rho^*} < 0$$

Proposition 7

To show that the sensitivity of aggregate asset risk to a change in liquidity requirements is higher for more levered banks, we need to show that $\frac{\partial^2 \rho^*}{\partial \xi \partial k} > 0$. We have that

$$\text{sign} \left(\frac{\partial^2 \rho^*}{\partial \xi \partial k} \right) = \text{sign} \left(1 + \frac{\partial^2 M_r^H}{\partial \xi \partial k} \right) \quad (3.34)$$

$$+ p_H \left(\frac{\partial M_r^H}{\partial k} L''(M_r^H) \frac{\partial M_r^H}{\partial \rho^*} + \frac{\partial^2 M_r^H}{\partial \rho^* \partial k} L'(M_r^H) \right) \quad (3.35)$$

$$- p_H \left(\frac{\partial M_0^L}{\partial k} L''(M_0^L) \frac{\partial M_0^L}{\partial \rho^*} + \frac{\partial^2 M_0^L}{\partial \rho^* \partial k} L'(M_0^L) \right) \quad (3.36)$$

It can then easily be verified that

$$\frac{\partial^2 \rho^*}{\partial \xi \partial k} > 0$$

Proposition 8

From equations 3.7 and 3.8, it is quite trivial to see that

$$\frac{\partial \rho}{\partial p_H} > 0$$

$$\frac{\partial \gamma}{\partial p_H} < 0$$

Chapter 4

Assessing Transit Rents (joint with Katrin Tinn)

Trading frictions due to inevitable transportation costs are fundamentally different from those due to rent extraction by transit countries. We propose a theoretical and empirical methodology to disentangle these two types of costs and assess the presence and global magnitude of a hold-up problem. We construct a new measure of distance based on a global network of the most likely trade routes. While transportation costs make all countries worse off, rent extraction benefits transit countries. Further, we show that in general equilibrium, countries that are neither landlocked nor transit countries bear a large share of the cost of distortions due to rent extraction. While free trade agreements with transit countries do not appear to mitigate the problem, customs unions do.

4.1 Introduction

Geographical location is important to international trade: countries that are closer to each other tend to trade more, and being landlocked is associated with less trade (see e.g., Frankel and Romer (1999), Rose (2004), and Overman, Redding, and Venables (2008)). Transportation costs that increase with distances and costs of air transport that are higher than those of sea and land transport have been invoked as partial explanations for this phenomenon.¹ Yet, trading frictions are not limited to solely technological obstacles of transportation. Adam Smith (1776) and Ronald Coase (1937) highlighted a potential hold-up problem, whereby intermediaries (e.g. ports) and perhaps transit countries have an opportunity and a strategic incentive to extract monopolistic or oligopolistic rents for letting goods pass through.²

Such rents to "the middlemen" are central to policy discussions on trade involving developing countries.³ Using micro-level data of trade by Ethiopian and Nigerian firms, Atkin and Donaldson (2015) find that a substantial part of the surplus generated by trade is captured by intermediaries. Furthermore, hold ups in freight are also argued to be a substantial obstacle to trade by developed countries. This observation has prompted discussions about the adoption of new technologies that can improve supply chain management.⁴

A geographically advantageous location may provide transit countries

¹Shipping a 70kg worth goods from Shanghai to London has been estimated to cost four times as much (and takes three times longer) than buying an airline ticket to a human of the same weight (see *Economist* 26 April 2018). See also Limão and Venables (2001) or Gaël Raballand, Kunth, and Auty (2005) for evidence that landlocked countries face higher transport costs.

²In contrast to the incomplete contracts literature which often focuses on the ex ante underinvestment implications of rent extraction (see Chemla and Milone, 2017), this international trade literature on the hold-up problem focuses on rent extraction rather than measuring the economic consequences of rent extraction.

³For example, Gael Raballand et al. (2012) explore delays in six African ports and argue that these delays serve the interests of public and private actors with market power. Relatedly Arvis, Gael Raballand, and Marteau (2007) argue that inefficient logistics may be a more important obstacle to trade for landlocked countries than infrastructure development. See also Djankov, Freund, and Pham (2010) and USAID (2004).

⁴For example, an article in *The Economist* 26 April 2018 argues that intermediaries involved in transfers between different means of transportation, custom clearances, insurance, and bureaucracy account for a fifth of logistics industry revenues, highlighting the promises of blockchain technology and some related initiatives in reducing these frictions.

with an opportunity to extract rents just like an advantageous position in the supply chain may enable firms to engage in rent seeking. If transit countries do extract rents, trade is affected by another layer of hold-up the magnitude of which depends on geographical locations. In particular, this hold up should affect countries that need to use a port in another country more than countries that have their own port. This differential country-level hold-up poses a potentially complex international coordination and political economy problem, which, unlike transportation technology or the reduction of intermediaries' monopolistic power, may not be easily solved with technological innovation (including innovation to supply chain management) or better competition regulation. At the same time, one may expect country level hold-ups to be historical problems that have been eliminated or alleviated by bilateral and multilateral trade agreements.⁵

The goal of this paper is to suggest a methodology to examine and assess the country-level hold-up problem aforementioned as well as its global impact. We start by developing a global general equilibrium model of trade that explicitly takes into account the fact that some trades need to go through a transit country due to geographical constraints. The global general equilibrium approach is essential as local trading frictions have a global impact via general equilibrium price effects. This was notably highlighted by Anderson and Van Wincoop (2003) who emphasize the importance of empirically capturing these price effects in gravity equations. We build on their framework and include the possibility that in addition to transport (and other "iceberg") trading costs, some countries can charge "transit fees" which differ from iceberg costs as they are a source of revenues for transit countries. Our model shows that there is a global deadweight loss that falls largely on the group of directly unaffected countries (like the UK, Japan, Australia, USA, or Mexico) that are in geographical locations that enable them to trade directly with most countries and are unlikely to obtain transit rents themselves. The reason is that while from the perspective of landlocked countries transport costs and hold-up frictions have a similar effect, typical transit countries ben-

⁵See e.g., Franck and Brownstone (1986) that discusses the Silk Road and tax collection on route. Also, Hirschman (1945) discusses trade policy as a tool for global influence.

efit from the country level hold up. In contrast, losses due to transport and other iceberg costs are likely to be more evenly spread across all countries.

Our theoretical exercise further suggests that in a world where the country level hold-up problem is empirically relevant, estimating gravity equations without considering transit rents may overstate the role of transport costs and socioeconomic ties when explaining why landlocked countries trade less and why neighboring countries trade more. We show that such pattern would be present also in a hypothetical world where all countries are identical apart from their location, and transport is costless. Furthermore, a cross-country structural gravity estimation approach that builds on Anderson and Van Wincoop (2003) and does not distinguish transit fees and iceberg-type trading costs leads to systematically biased estimates of global price effects, frequently called "multilateral resistance terms".

Empirically, distinguishing transit rents from transportation costs requires a good measure of distance. We propose a novel and flexible measure of distance: we construct a world-wide network of theoretically possible trade routes between all main cities that considers the location of major ports and differentiates between land and sea distances. Our algorithm then generates the most likely trade routes solely based on geography, the relative cost of sea and land transport, and potential fixed costs associated with loading and unloading goods. An advantage of that approach is that it enables us to identify the most likely transit countries based on geography alone, and it is flexible enough to easily consider alternative relative and fixed costs, some of which we have considered in our robustness checks. An obvious benefit of using this measure compared to the traditional great-circle distance measure between the main cities is that the great-circle distance tends to minimize long trade routes, potentially resulting in an underestimation of bilateral transportation costs, which is particularly relevant for trade with landlocked countries. The great-circle distance cannot also provide information regarding the type of transportation used (land or sea)

Using 1993-2016 trade data, we show that the hold-up problem exists and is statistically and economically significant. On average trades that are likely to go through a transit country could be increased by approximately

28% were they not held up. Even though bilateral trades that are exposed to transit rents often involve a landlocked country, it is important to emphasize that the country-level hold up differs from a country's status on being landlocked. Being landlocked is a static, country-specific characteristic that is captured by (time varying) fixed effects included in our analysis. If lower trade by these countries would relate to specific characteristics of landlocked countries, we would not find significant results. In contrast, whether or not a particular trade is potentially held up is a bilateral and trade-specific problem: for example the trade between a landlocked country and its neighbor is not held up, while its trades with further away countries is. Furthermore, as land transport is not prohibitively expensive trade between two coastal countries may benefit from going through a transit country (e.g., trade between Slovenia and Germany may benefit from going through Austria) and could be held up as a result. A landlocked country itself could be a transit country that can charge rents in some bilateral trade relations (e.g., Austria in the case of trade between Slovenia and Germany). There are indeed numerous cases where landlocked countries are also transit countries.

Our estimation strategy does not aim to identify the particular method of rent extraction the transit country uses. Obvious examples of possible methods include tariffs, road tolls, and different port handling costs for good in transit and goods that are part of exports from or imports to the country of the port. Furthermore, the monopolistic/oligopolistic power of ports that intermediate large volumes of trade in transit alone enables them to extract rents: even if the port charges similar rents for all firms, both profits and the income of its employees remain within the transit country's GDP and there is still a de facto wealth transfer between the held up country and the transit country. Hence transit countries may have a limited interest in reducing the market power of ports. Relatedly, it is worth emphasizing that that the hold-up problem we identify is the differential one and would not capture the cost at ports that falls identically on goods traded by domestic firms and goods in transit.

We further consider a number of robustness checks, including differences across continents. The hold-up problem appears to be most severe in Africa

and least severe in Europe, but is present in all continents. As trade between many European countries is facilitated by European Union and free trade agreements and customs unions also exists elsewhere, a natural question is whether the hold-up problem is mitigated, or perhaps eliminated, by these agreements. We find that while free trade agreements between the hold up and the transit countries do not appear to mitigate the problem, customs unions reduce the hold-up friction. Indeed, central features of a customs union are that countries charge identical import duties to each other and typically allow free trade between themselves. Nevertheless, customs unions do not appear to eliminate the country-level hold-up friction, which suggests that transit rents may indeed include more indirect features, e.g., differential costs at ports, monopolistic power of ports, road tolls etc.

After establishing the hold up and indirectly assessing transit rents, we use our model and data to simulate the world without the hold-up problem and transit rents. We find that if these rents were to be eliminated, the global gains would be around 350 billion 2017 USD per year.⁶ These gains are not evenly distributed. Absent transit rents, the group of likely transit countries would lose approximately USD 350 billion, but landlocked countries would gain USD 500 billion. The group of countries that are *not* directly affected (countries such as the UK and the USA), would gain around USD 200 billion annually. We repeat this exercise considering that transit countries in customs unions charge lower transit fees when intermediating trades that involve other countries in the same customs union. The global gains and losses are qualitatively similar albeit quantitatively smaller. The latter is to be expected as customs unions include many relatively richer countries. We contrast this prediction to a simulation of the world without transit rents to one where there is a reduction of transport costs leading the exactly the same global gain. In that simulation all groups of countries gains, and importantly the gains of directly unaffected countries is much smaller. This highlights that mitigating country level hold up would be of benefit to a wide set of countries and not just landlocked countries, but would not be in the interest

⁶Estimates vary if one considers different elasticities of substitution, or the beneficial effect of customs unions.

of transit countries. Yet, reducing this problem does not require technological innovation, and could at least theoretically be tackled by limiting the market power of transit countries and ports while compensating for the losses of transit countries.

The paper builds on, and contributes to, the gravity equations literature (see e.g., Head and Mayer (2014) and Costinot and Andrés (2014) for elaborate overviews including the relationship between theoretical models of trade and empirical methodologies). As our question is about global general equilibrium effects of country level hold up and transit rents, and micro level data does not exist in standardized format at the global scale, our theoretical setting builds on Armington (1969) and Anderson and Van Wincoop (2003) where country level production is exogenous. It also provides a simple framework to highlight the main distortions generated by transit rents. Namely, even with exogenous production, transit countries wealth in the presence of transit rents is endogenous. We show that global price indexes are affected by transit rents, and multilateral resistance terms associated with importing and exporting country are not symmetric, even if physical trading costs are symmetric. Furthermore, as transit rents are not observable, but enter in measured GDP, our method of identifying transit countries enables us to provide preliminary assessment of global welfare costs of country level holdup.⁷

While the estimated magnitude of welfare costs is model specific, the same qualitative predictions would also emerge in other settings. Namely, the gravity equation that our model generates has a standard form, and is well known to emerge in many international trade models that explicitly consider production choices, different forms of production between firms and firm level heterogeneity (see e.g. Krugman (1980), Eaton and Kortum (2002), Bernard et al. (2003), Chaney (2008), Arkolakis et al. (2008), and Arkolakis (2010)) In fact, welfare gains obtained by reducing trade barriers are known to be

⁷It should be noted that we use the "best guess" based on geography to identify transit countries, and may at times attribute transit rents to a wrong country. Such errors are likely to average out across countries, and are unlikely to bias global and country group specific estimates. However, our method or data should *not* be used as a measure of transit rents for a specific country.

higher if one considers imperfect competition between firms (see e.g., Table 4.1 in Costinot and Andrés (2014)). Hence, the welfare gains associated with the transit rents we estimate are likely lower than those that would emerge in settings that incorporate realistic firm level effects. For example, introducing the possibility of transit rents to M. J. Melitz (2003)'s setting and considering the effect of a reduction in the hold-up problem will likely lead to larger global gains due an increase in productivity via a country-level endogenous shift toward more productive firms. Such additional effects could be assessed when more micro-level data is available globally. At the same time it should be noted that unlike the estimation on welfare, our estimates of the presence of hold up and its effect on log-export do not rely on a particular model, as they are based on a gravity equation that holds more generally, as argued above. This argument also benefit from results by Feenstra (2004) and Redding and Venables (2004) who show that using (time varying) exporter and importer fixed effects produces unbiased estimates of trade frictions based on gravity equations that emerge in different settings.

Motivated by historical examples such as the silk road, the possibility that transit countries charge monopolistic rents features in a number of theoretical papers that consider trade through a chain of markets which can impose taxes and tolls (see e.g. Karni and Chakrabarti (1997), Feinberg and Kamien (2001), Gardner et al. (2002), and Miyagiwa (2009)). These focus on game theoretical questions of rent extraction by multiple intermediaries and do not consider general equilibrium pricing and welfare effects on which we focus here. Related research analyzes and estimates the importance of the market power of shipping industry: Hummels, Lugovskyy, and Skiba (2009) highlight the role of markups and estimate that the gains from eliminating market power in the shipping industry in the US and Latin America would lead to noticeable gains, and particularly so Latin America. These findings are complementary to ours as we emphasize and estimate the gains from eliminating a differential hold-up problem which generates additional costs to port level mark-ups. The benefits of eliminating both types of hold-up frictions would likely lead to gains aggregating these two effects.

Our paper also relates to the literature discussing the measurement of

the distance and the possible bias generated by the commonly used great-circle distance measure. Many proposed measures aim to improve within-country distance measures, e.g., Atkin and Donaldson (2015) use Google maps data on road distances in Ethiopia and Nigeria, Head and Mayer (2010) advocate district to district distances and review other related papers that highlight weaknesses using great circle distances between main cities. In specific contexts, it is also possible to use survey data on transport costs (see e.g., G. J. R. F. Raballand and Teravaninthorn (2008)). In this paper, we are less concerned about within country distances as our question is about cross country trade, and any mismeasurement within a country is likely to be a fixed characteristic that is largely captured by fixed effects. Instead, our distance measure focuses on location of ports and enables us to draw transportation networks that depend on relative costs of sea and land, which can vary over time.

Our paper also contributes to the literature on the role of trade agreements (see Maggi (2014) for a review). Our focus is somewhat different here, as we are interested in the relevance of these agreements for mitigating the hold-up problem, and hence the role of trade agreement with the transit country. Customs unions and free trade agreements have a primary and direct relevance in the context of bilateral trade beyond including bilateral agreements as a control.

4.2 A model of global trade with transit rents

4.2.1 The setting

We incorporate transit rents in a model of global trade in the spirit of Armington (1969) and Anderson and Van Wincoop (2003). That is, each country j produces a distinct variety and its production is exogenously given and denoted with Y_j . The benefits from trade arise because the representative consumer in each country has preferences for different varieties, i.e., the rep-

representative consumer in country j solves

$$\max_{\{c_{ij}\}} U_j \equiv \left(\sum_i \beta_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \text{ s.t. } \sum_i p_{ij} c_{ij} = W_j \quad (4.1)$$

where $\sigma > 1$ is the elasticity of substitution, c_{ij} is the consumption of a good produced in country i in country j , and β_i is a taste parameter that captures the subjective preferences regarding country i 's good, p_{ij} denotes the price of a good produced in country i in country j and W_j is the nominal wealth of consumers in country j . The budget constraint equates the nominal wealth with spending of country j 's consumer on goods produced in each country j . As in many related papers, solving (4.1) leads to the following demand function

$$c_{ij} = (p_{ij})^{-\sigma} W_j \frac{(\beta_i)^{1-\sigma}}{P_j^{1-\sigma}}, \quad (4.2)$$

where a country specific price index is defined as

$$P_j \equiv \left(\sum_i (\beta_i p_{ij})^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \quad (4.3)$$

In nominal terms (in importing country prices), the export from i to j is

$$X_{ij} = p_{ij} c_{ij} = W_j \frac{(\beta_i p_{ij})^{1-\sigma}}{P_j^{1-\sigma}} \quad (4.4)$$

Due to physical trading costs and transit rents, consumers in different countries may pay different prices for goods produced in country i . However, goods exported to different countries are perfect substitutes at the producing country. We denote the producer price in country i with p_i . As standard, we allow for iceberg trading costs: in order to deliver c_{ij} units from i to j , country i must produce $\tau_{ij} c_{ij}$ units, where $\tau_{ij} \geq 1$. We refer to τ_{ij} as "the transportation cost", but it can include other costs associated with bilateral trade (additional controls which we later include in our empirical analysis).

Our main innovation is to allow transit rents in addition to these standard exogenous trading costs. The final price that country consumers in country

j pay for country i 's good is

$$\begin{aligned} p_{ij} &= \tilde{\mathbf{1}}_{mij} \varphi_{mij} \tau_{ij} p_i + (1 - \tilde{\mathbf{1}}_{mij}) \tau_{ij} p \\ &= \tau_{ij} p_i (1 + \tilde{\mathbf{1}}_{mij} (\varphi_{mij} - 1)) \end{aligned} \quad (4.5)$$

where p_i is the producer price in country i ; $\tilde{\mathbf{1}}_{mij}$ is an indicator function, where $\tilde{\mathbf{1}}_{mij} = 1$ if due to geography, the trade between country i and j must go through a transit country m (a country in "middle"), and is zero otherwise. The variable $\varphi_{mij} \geq 1$ measures the additional cost imposed by country m on bilateral trade between countries i and j ($i, j \neq m$). If i and j trade directly (e.g., they are neighbors or they both have sea access and the best trade route between them is via the sea⁸) then $\tilde{\mathbf{1}}_{mij} = 0$. To shorten the notation, we will also use

$$\tilde{\varphi}_{mij} \equiv 1 + \tilde{\mathbf{1}}_{mij} (\varphi_{mij} - 1) \quad (4.6)$$

such that $p_{ij} = \tau_{ij} p_i \tilde{\varphi}_{mij}$. In the presence of iceberg cost, the market clearing condition within each country is

$$Y_i = \sum_j \tau_{ij} c_{ij}, \quad (4.7)$$

i.e., production of country i 's good must equal the global consumption of country i 's good in real terms.

Finally, it is important to emphasize that the nominal wealth of consumers in a country depends on whether, or not, the country is a transit country. The wealth of country m that is a transit country is

$$W_m = p_m Y_m + T_m, \quad (4.8)$$

where T_m is the total a transit fee that country m obtains. The total transit fee is

$$T_m = \sum_j \sum_i \tilde{\mathbf{1}}_{mij} (\varphi_{mij} - 1) \tau_{ij} p_i c_{ij}.$$

If a country j is never a transit country for any trading relations (e.g., USA),

⁸Or via a large lake as it is the case of countries around the Caspian Sea

then $T_j = 0 \implies W_j = p_j Y_j$. Note that the measured wealth/nominal income (empirically measured by GDP) of a country includes any transit rents the country receives and in the case of transit countries we need to rely on the model to distinguish transit rents and local production. Finally, the global nominal wealth must equal nominal income from production in all countries and total income from transit fees charged by all countries, i.e.,

$$\sum_i W_i = \sum_i p_i Y_i + \sum_i T_i.$$

4.2.2 Gravity equation and multilateral resistance terms

From (4.5) and (4.6), we have $p_{ij} = \tilde{\varphi}_{mij} p_i \tau_{ij}$. Hence, using market clearing (4.7) and export (4.4), we obtain

$$(\beta_i p_i)^{1-\sigma} = \frac{p_i Y_i}{\Pi_i^{1-\sigma}}, \quad (4.9)$$

where

$$\Pi_i \equiv \left(\sum_j \frac{1}{\tilde{\varphi}_{mij}} W_j \frac{(\tilde{\varphi}_{mij} \tau_{ij})^{1-\sigma}}{P_j^{1-\sigma}} \right)^{\frac{1}{1-\sigma}}, \quad (4.10)$$

is the multilateral resistance term associated with the exporting country. We can then use (4.4), (4.9) and (4.8) to obtain the following gravity equation

$$X_{ij} = (\tau_{ij} \tilde{\varphi}_{mij})^{1-\sigma} \frac{W_j \cdot p_i Y_i}{P_j^{1-\sigma} \Pi_i^{1-\sigma}} = (\tau_{ij} \tilde{\varphi}_{mij})^{1-\sigma} \frac{W_j \cdot (W_i - T_i)}{P_j^{1-\sigma} \Pi_i^{1-\sigma}}, \quad (4.11)$$

where the price index (multilateral resistance term associated with the importer)

$$P_j = \left(\sum_i (p_i Y_i) \frac{(\tilde{\varphi}_{mij} \tau_{ij})^{1-\sigma}}{\Pi_i^{1-\sigma}} \right)^{\frac{1}{1-\sigma}} = \left(\sum_i (W_i - T_i) \frac{(\tilde{\varphi}_{mij} \tau_{ij})^{1-\sigma}}{\Pi_i^{1-\sigma}} \right)^{\frac{1}{1-\sigma}} \quad (4.12)$$

From (4.11), it is clear that hold up and transit rents reduce bilateral exports via three channels. First the additional costs due to transit rents, $\tilde{\varphi}_{mij}$, reduce trade similarly to transportation costs. Second, in the exporting country

i is a transit country that receives fee income ($T_i > 0$), it trades less at a given level of wealth because it produces less. Third the multilateral resistance terms are further indirectly affected by transit fees and rents anywhere else in the world. The latter effect is also present for bilateral trades that are not held up ($\tilde{\varphi}_{mij} = 1$) or do not involve a transit country. From (4.10) and (4.12) transit rents and hold up fees imply higher price indices P_j and Π_i (and lower $P_j^{1-\sigma}\Pi_i^{1-\sigma}$). This in turn implies that trade between countries that are not directly affected, such as two non-transit neighbors, is higher.

Anderson and Van Wincoop (2003) consider trade frictions that are symmetric, i.e., $\tau_{ij} = \tau_{ji}$, which imply symmetric multilateral resistance terms, facilitating their structural estimation. In a setting with transit rents, the symmetry of trade frictions, i.e., $\tau_{ij} = \tau_{ji}$ and $\tilde{\varphi}_{mij} = \tilde{\varphi}_{mji}$, no longer implies the symmetry of the multilateral resistance terms. Hence, $P_j \neq \Pi_i$ and a symmetric approach cannot be used.

Notice also that from (4.11)

$$\frac{X_{ij}}{W_j W_i} = \frac{(\tau_{ij} \tilde{\varphi}_{mij})^{1-\sigma} (W_i - T_i)}{P_j^{1-\sigma} \Pi_i^{1-\sigma} W_i},$$

where the left hand side is observable by data. If the correct model includes transit rents but the econometrician only considers iceberg costs, he would consider the left hand side to be $\frac{(\tau_{ij})^{1-\sigma}}{\check{P}_j^{1-\sigma} \check{\Pi}_i^{1-\sigma}}$, where \check{P}_j and $\check{\Pi}_i$ are multilateral resistance terms under the incorrect model. The estimated $\check{P}_j^{1-\sigma} \check{\Pi}_i^{1-\sigma} = P_j^{1-\sigma} \Pi_i^{1-\sigma} \left((\tilde{\varphi}_{mij})^{1-\sigma} \frac{(W_i - T_i)}{W_i} \right)^{-1}$. Since $(\tilde{\varphi}_{mij})^{1-\sigma} < 1$ for bilateral trades that are held up and $\frac{(W_i - T_i)}{W_i} < 1$ when the exporter is a transit country, it follows that in these cases $\check{P}_j^{1-\sigma} \check{\Pi}_i^{1-\sigma} < P_j^{1-\sigma} \Pi_i^{1-\sigma}$. Hence, the econometrician would systematically overestimate the multilateral resistance terms for such trading pairs. Consequently, from the analysis of (4.11), the econometrician may attribute too much importance to the distance (less trade with far away countries, more trade with neighbors) and to cultural and common characteristics of nearby countries (see also Section 4.2.5).

Finally, notice that terms $\frac{W_j}{P_j^{1-\sigma}}$ and $\frac{(W_i - T_i)}{\Pi_i^{1-\sigma}}$ in (4.11) are country specific. As highlighted in Feenstra (2004), one can obtain unbiased estimates of τ_{ij} and $\tilde{\varphi}_{mij}$ by considering logs and including exporter and importer fixed

effects. Section 4.5 will discuss this further.

4.2.3 Welfare

Using the budget constraint, (4.2) and (4.3) in (4.1) we derive the indirect utility of the representative consumer in country j as

$$U_j = \frac{W_j}{P_j}$$

In order to assess the global welfare gains and losses in units that can be interpreted, it is useful to transform this utility to a *money metric* indirect utility function constructed via the expenditure function $e(P_j, U_j) = P_j U_j$ (see Section 3.I in Mas-Colell, Whinston and Green (1995)). We adopt equivalent variation as our measure of welfare change that results from moving from the existing world to an alternative hypothetical world that consider some change in trading costs (such as elimination of transit fees), i.e., the welfare change in country j is

$$EV_j(P_j, \Pi P_j, W_j, \Pi W_j) = e(P_j, \Pi U_j) - e(P_j, U_j) = P_j \Pi U_j - P_j U_j = \frac{\Pi W_j}{\Pi P_j} P_j - W_j, \quad (4.13)$$

where P_j and W_j is the observed prices and wealth, and ΠP_j and ΠW_j are the prices and wealth in the alternative world. Equivalent variation reflects the monetary amount that the representative consumer in country j would be indifferent about accepting (or paying) instead of living in such alternative world. As our data is in US dollars the equivalent variation in each country is also measured in US dollars. It is then also straightforward to measure utility changes globally and within a specific group of countries (e.g., landlocked) in US dollars by the sum

$$\sum_{j \in \{group\}} EV_j(P_j, \Pi P_j, W_j, \Pi W_j).$$

4.2.4 Monopolistic fees charged

The fees collected from different trades by different intermediates are additive and can be thus analyzed separately. Suppose that the export from country i to j needs to go through country m . The income from the transit fee that country m obtains from this trade is

$$Fee_{ij}^m = (\tilde{\varphi}_{mij} - 1) \tau_{ij} p_i c_{ij} = \frac{(\tilde{\varphi}_{mij} - 1)}{\tilde{\varphi}_{mij}} X_{ij},$$

where we used $p_{ij} = \varphi_{ij}^m \tau_{ij} p_i$ and export $X_{ij} \equiv p_{ij} c_{ij}$. A monopolistic transit country takes global prices and wealth levels as given, and internalizes the fact that fees affect trade flows, i.e. it chooses $\tilde{\varphi}_{mij}$ to maximize Fee_{ij}^m subject to export given by (4.11). Hence,

$$Fee_{ij}^m = \frac{(\tilde{\varphi}_{mij} - 1)}{\tilde{\varphi}_{mij}^\sigma} (\tau_{ij})^{1-\sigma} \frac{W_j \cdot (W_i - T_i)}{P_j^{1-\sigma} \Pi_i^{1-\sigma}}$$

and the optimal transit fee is

$$\tilde{\varphi}_{mij} = \frac{\sigma}{\sigma - 1}, \quad (4.14)$$

i.e., a monopolistic transit country charges a constant mark-up. This relationship is useful as it provides a relationship between hold up fees and the elasticity of substitution σ that is not directly observable.

One could further speculate that transit countries may have less market power, e.g., the good could take an alternative route or there is an agreement between countries that limits the transit country's limit power. Under such alternative scenarios, the transit fees could be expressed as $\tilde{\varphi}_{mij} = \chi \frac{\sigma}{\sigma - 1}$, where $\chi \in [0, 1]$, where lower χ indicates less market power and a lower mark-up.

4.2.5 Example with four countries

To highlight the main differences between transport (and other iceberg-type) costs and transit fees consider a greatly simplified world depicted on Figure

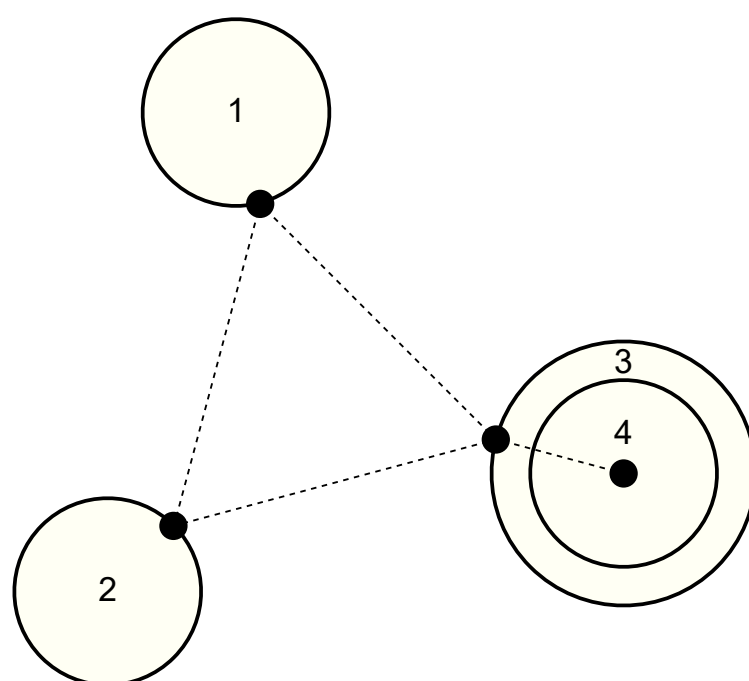


Figure 4.1: Topology of the four countries example.

4.1. Countries 1 and 2 are directly unaffected countries (islands), Country 3 is a transit country and Country 4 is a landlocked country, whose trade with Countries 1 and 2 is help-up and whose trade with Country 3 is not held-up. The circles mark the centers of economic activity in each country, and the dotted lines represent shortest transport routes. Furthermore, trade between Countries 1,2,3 takes place via sea transport, while in order to trade with any other country, goods from Country 4 must use road transportation to reach the main city/port in Country 3. For the sake of clarity, assume that all countries $j = \{1, 2, 3, 4\}$ are equally productive, i.e., $Y_j = Y$ for all j , and goods from all countries are equally valued $\beta_j = 1$. We normalize the price of the good produced in Country 1, $p_1 = 1$, and assume that the elasticity of substitution is $\sigma = 5$.

As a benchmark, assume a frictionless world where Country 3 does not charge transit rents and both land and sea transport is costless. Since all countries are identical under this benchmark economy, it follows from (4.1), (4.2), (4.3), and (4.4) that prices and price indexes are the same in all countries, $p_j = 1$ and $P_j = 4^{\frac{1}{1-\sigma}} = \frac{\sqrt{2}}{2}$, and the nominal income in all countries is the same, $W_j = p_j Y = Y$ for all j . Furthermore, the representative consumer in each country consumes domestic and foreign goods from each other country in equal proportions, i.e., $c_{ij} = \frac{Y}{4}$ for all $i, j = \{1, 2, 3, 4\}$ and $\frac{\text{import}}{\text{GDP}} = \frac{W_j - p_j c_{jj}}{W_j} = 75\%$. Weighting all countries equally, the global indirect utility is $U_w^B = \sum_j \frac{W_j^B}{P_j^B} = \frac{8Y}{\sqrt{2}}$.

We then compare three stylized cases of different frictions, such that the global utility under all these cases, i.e., $U_w = \sum_j \frac{W_j}{P_j}$ is constant. We consider frictions lead to a global utility loss $U_w/U_w^B - 1 = -1.3\%$.

Case I is the world where transport is costless, and Country 3 charges monopolistic transit rents whenever Countries 1 or 2 trade with Country 4. From (4.14) Country 3 charges the same the same rents $\varphi = \frac{\sigma}{\sigma-1} = 1.25$ from all trades that pass through. Case II is the world where there the sea transport remains costless and there are no transit rents, but land transport is subject to an iceberg cost τ_L . This implies that trade with Country 4 is costly, i.e., $\tau_{i4} = \tau_{4i} = \tau_L > 1$ for $i = \{1, 2, 3\}$, and the trade between 1-2-3

is costless. We use $\tau_L = 1.0375$ to keep the global utility loss constant. Case III considers the same costs for land and sea transport, and no transit rents, i.e., there is an iceberg cost $\tau_{i4} = \tau_{4i} = \tau^2 > 1$ for trade between 4 and $i = \{1, 2\}$, and an iceberg cost cost is τ for all other bilateral trades. This last case corresponds to a world where transport costs can be well captured by the great circle distance. We use $\tau = 1.013$ to keep the global utility loss constant.

Table 4.1 presents percentage utility gains and losses ($\% \Delta U$), and percentage changes in wealth ($\% \Delta W$), price indexes ($\% \Delta$ in P) and producer prices ($\% \Delta p$) relative to the frictionless benchmark. It also reports the import/GDP ratio $\frac{imp_j}{W_j} \equiv \frac{W_j - p_j c_{jj}}{W_j}$.

Table 4.1 highlights that while different transportation costs reduce utilities in all countries, the utility losses are spread across different countries. While landlocked countries are affected more, all other countries have lower utility as well due to price effects (the producer prices fall and the price indexes that affect the consumption basket increase). These effects are well known and can partially explain the observed lower trade by landlocked countries and as well as somewhat less trade by all countries (i.e., import to GDP ratios are lower than in frictionless world as the consumption baskets are tilted towards the consumption of domestic goods). As it is to be expected, the losses of a landlocked country are relatively bigger when land transport is noticeably more expensive than sea transport.

Transit rents have a very different impact on the distribution of losses across countries. These rents benefit transit countries, and noticeable losses are incurred not just by the landlocked country 4, but also by the seemingly less affected "islands" 1 and 2. While the specific values are specific to the example, the fact that a large part of losses falls on such countries is a general pattern. There is a global deadweight loss, and the transit country 3 gains, which means that the combined welfare losses of Countries 1, 2, and 3 must be greater than the global loss. Furthermore, notice that the transit country 3 gains via three channels: first, it gets an additional income from rents; second, as it is not held up itself, it can buy goods from all other countries more cheaply (its price index falls); third, it sells its own production at a

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
$\% \Delta U$	-11.7	+15.6	-4.5
$\% \Delta W$	-3.7	+15.5	na
$\% \Delta P$	+9.0	-0.1	+4.5
$\% \Delta p$	-3.8	+4.2	na
$\% \frac{imp}{W}$	57	76	70

(a) I Transit rents

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
$\% \Delta U$	-3.6	-0.5	-0.5
$\% \Delta W$	-1.4	na	na
$\% \Delta P$	+0.6	+0.6	+2.3
$\% \Delta p$	-1.4	na	na
$\% \frac{imp}{W}$	71	74	74

(b) II: Costly land transport

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
$\% \Delta U$	-1.6	-0.9	-1.3
$\% \Delta W$	-0.03	-0.04	na
$\% \Delta P$	+1.6	+1.0	+1.3
$\% \Delta p$	-0.03	-0.04	na
$\% \frac{imp}{W}$	73	74	74

(c) III: Costly land and sea transport

Table 4.1: Changes in Utility, price indices and production prices

higher price everywhere. It also uses its extra wealth to tilt its consumption towards foreign goods (its import/GDP ratio is higher than in the frictionless benchmark). This explains why "islands" 1 and 2 must bear a substantial part of utility losses in general equilibrium: not only it is expensive for them to buy goods produced in Country 4, but also Country 3's goods are more expensive relative to their own production (numeraire). All this increases the price of the consumption basket in Countries 1 and 2 and sub-optimally tilts consumption towards their own good. As expected, transit rents have a similar negative price effect on the landlocked country and the relative price of its own production is lower. Consequently, its consumption basket is even more heavily tilted towards own good.

Table 4.2 further reports the percentage changes of quantities consumed demand (i.e., Hicksian demand) compared to the benchmark under these three scenarios. The columns reflect the compositions of consumption baskets in a given country. The rows reflect the origin of the goods consumed.

Table 4.2 shows that while trading frictions generally lead to sub-optimally high consumption of domestic goods, the patterns of trade between countries are quite different. In particular, transit rents encourage more consumption of neighboring country's goods and thus trading between neighboring countries. Note that under scenario I, countries 1 and 2 trade noticeably more with each other, and also landlocked countries trade noticeably more with the transit country. Such neighboring trade patterns are not by far as striking when we consider different types of transport costs only.

Consider an econometrician who observes data generated in a world with transit frictions (case I), but considers an empirical setting based on iceberg costs and the great circle distance (case III). The econometrician will likely conclude that the noticeably higher trade between neighboring countries must be due to unobservable common tastes, which we assumed not to be the case here. Furthermore, the econometrician is also likely to attribute greater wealth in the transit country and lower wealth in the landlocked country to be due to differences in productivity, which we also assumed not to be the case. While common tastes and productivity differences are likely to be present in the real world data, this example illustrates that the estimates

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
Domestic	+64.7	+13.1	+20.1
Import from landlocked	na	+39.6	-55.5
Import from transit	+10.5	na	-2.1
Import from other	-55.5	+15.1	+20.1

(a) I Transit rents

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
Domestic	+16	+2.2	+2.2
Import from landlocked	na	-8.8	-8.8
Import from transit	-10	na	+2.2
Import from other	-10	+2.2	+2.2

(b) II: Costly land transport

	Country 4 Landlocked	Country 3 Transit	Country 1&2 Other
Domestic	+6.7	+5.1	+5.2
Import from landlocked	na	-2.4	-1.5
Import from transit	-0.3	na	-7.4
Import from other	-6.4	-2.5	-1.3

(c) III: Costly land and sea transport

Table 4.2: Percentage changes in consumption demands

of these effects may be biased and overestimated when one ignoring transit rents.

4.3 Global Trade Network

An important contribution of our paper is to provide an accurate distance measure between trading countries. Gravity equations are typically estimated using the great-circle distance which provides at best a crude approximation of how far apart countries are from each other, underestimates long trading distances and does not provide any information about the type of trade route used. To properly investigate the holdup problem, one needs to be able to carefully analyze trade paths between any trading countries, and determine which are the most likely transit countries. Moreover, disentangling the hold-up friction from transport costs needs a better understanding of the trade path in terms of land or sea transportation.

Using a worldwide network of trade routes, we are able to overcome the shortcomings of the great-circle distance, arguably providing much more accurate estimations of the real trade distances. We generate valuable information as we can distinguish land and sea transportation as well as determine the most likely transit countries. The trade routes network is constructed as follows. We first create a network of all shipping lanes using data from the Oak Ridge National Laboratory.⁹ With port information from the World Port Source website¹⁰, we connect the shipping lanes to all ports with container liner services.¹¹ We then add all the countries' main cities¹² to their countries' respective ports and to the main cities of the neighbors countries using great circle distance.¹³ This procedure returns a complete worldwide

⁹See <http://oceanids.geoiq.grida.no/overlays/25>.

¹⁰See <http://www.worldportsource.com/>. World Port Source provides the location of ports worldwide as well as their uses and sizes.

¹¹If a country does not have any port with container liner service, we use its biggest ports.

¹²The main cities are provided by CEPII, and extended for over 20 countries.

¹³It is possible to create a graph with all real land routes, but it adds a lot of complexity and does not substantially increase accuracy. We aim for simplicity while accurately capturing trade routes.

graph of all land routes and shipping lanes between all countries.¹⁴

Using this graph, any trade route between any two countries can be approximated. The simplest way to do so is to find the shortest route between any two countries main cities. Doing so has several shortcomings, as realistic trade routes are not the ones with the shortest path. One needs to account for the facts that transportation costs differ between land and sea, and that it would be too costly to load and unload goods at several ports along the way. We obtain realistic trade routes by taking these two constraints into account. Instead of looking for the shortest paths in terms of distance, we look for the cheapest paths and assign differentiated costs between land routes and shipping lanes, as well as fixed cost for any loading/unloading of goods in ports. Land transportation must be expensive enough for long trading routes to be relying on shipping lanes while shorter trading routes are more likely to be using truck or rail shipments, as in Europe. In our main specification, we assume that land routes are 7 times more costly than shipping lanes.¹⁵ These calibrations seem to be consistent with real estimations of transportation costs. According to the U.S. Department of Energy and the U.S Department of Transportation ¹⁶ , U.S domestic shipping is estimated to be roughly between 7 to 12 times more energy efficient than trucks. To be confident that our methodology provides sensible trade routes, every single trade path pattern has been checked to start with road transportation followed by possible shipping, and ending by road. Figure 4.2 and 4.3 illustrate the predicted trade routes between the United Kingdom and the Central African Republic, Angola and India, France and Hungary as well as Slovenia and Germany.

Given the predicted trade paths transit countries between any two trade partners can be determined by analyzing countries through which each trade path passes. In our subsequent analysis of the impact of transit rents on

¹⁴Land routes (road and rail networks) being approximated by great-circle distances between neighbors.

¹⁵ We have considered other cost differentials and the results are robust to variations of this parameter.

¹⁶See U.S. Department of Energy, *Transportation Energy Data Book*, 2008. and U.S. Department of Transportation, *National Transportation Statistics*, 2009.



(a) United Kingdom – Central African Republic



(b) Angola – India

Figure 4.2: Predicted trade routes for long distances.



(a) France – Hungary



(b) Slovenia – Germany

Figure 4.3: Predicted trade routes for short distances.

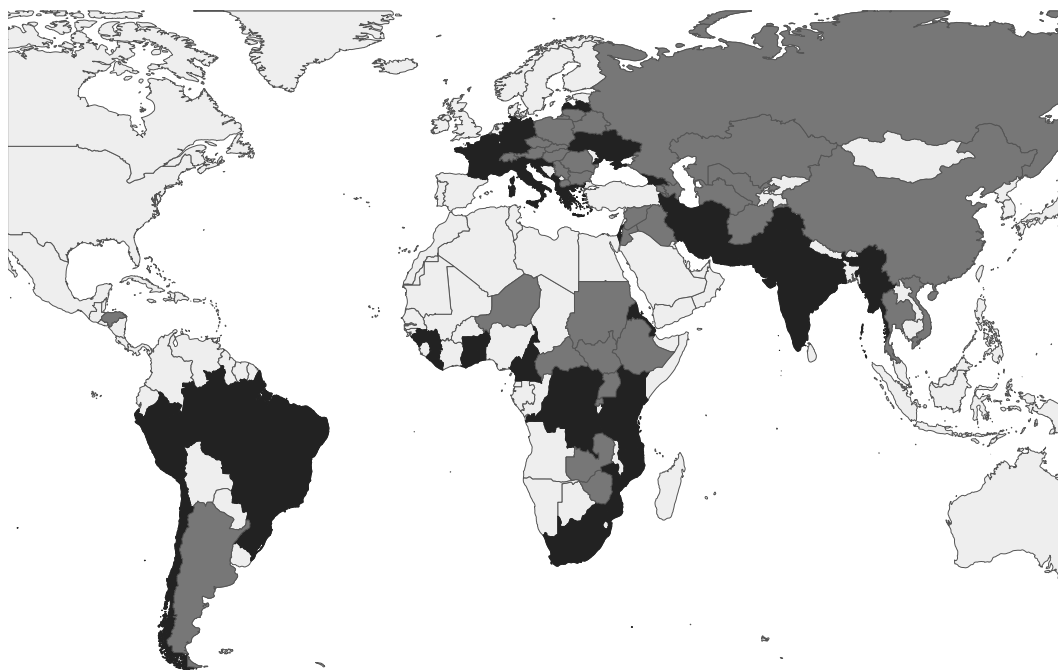


Figure 4.4: Transit Countries

welfare, we seek to differentiate transit countries and countries that are held up. However, depending on the trade partners considered, some countries may sometimes be transit countries or held up countries. Figure 4.4 presents the map of transit countries as identified by the global network of trades – countries in gray and black. In our welfare analysis we use the following methodology to define a clear non-overlapping set of transit and held up countries. From all the possible trade routes resulting from the global network, we classify as transit countries those that are more often transit than held up countries. These countries are the ones in black on the map.

[ADD ROAD COST REFERENCES IN THE BIBLIOGRAPHY]

4.4 Data and empirical strategy

We construct an extensive dataset of international trade. Trade statistics come from the IMF DOTS database¹⁷ with values expressed in millions of

¹⁷IMF Direction of Trade Statistics database.

constant 1982-1983 US dollars. We symmetrize trading amounts by averaging imports and exports for each country pair. If only one partner country reports the value of exports (imports), it is considered as being the value of imports (exports) for the other country. This procedure ensures the use of all available trade information, mitigates reporting errors, and minimizes the issue of missing data. We extend the trade statistics to include country pairs information. Main geographic information as well as languages and colonies are from the CEPII¹⁸ and manually extended to include all countries in our dataset. WTO memberships and regional trade agreements are from the WTO.¹⁹ They include all bilateral and multilateral trade agreements as well as customs unions. Currency unions are taken from various sources²⁰, and GSP (General Scheme of Preferences) information is from Andrew Rose²¹ dataset and updated for later years. All control variables are described in appendix 4.6. To have a large and consistent number of countries, notably countries from eastern Europe, our final dataset spans from 1993 to 2014 and includes trades between up to 200 countries.²² Because our dataset covers substantially more landlocked countries from 1993, most of our analysis is based on the 1993-2016 time period.

Our empirical strategy is twofold. First we aim to provide empirical evidence of the existence of transit rents and their impact on trade. Second we wish to estimate the economic significance and the welfare losses associated.

In order to test for the existence of the holdup problem, we use a regression approach and estimate the logarithmic form of the gravity equation (4.11).

$$\ln X_{ij} = (1 - \sigma) (\ln \tilde{\varphi}_{mij} + \ln \tau_{ij}) + \ln \left(\frac{W_j}{P_j^{1-\sigma}} \right) + \ln \left(\frac{W_i - T_i}{\Pi_i^{1-\sigma}} \right) \quad (4.15)$$

¹⁸Centre d'Etudes Prospectives et d'Informations Internationales.

¹⁹See https://www.wto.org/english/thewto_e/whatis_e/tif_e/org6_e.htm for WTO memberships and <http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx> for regional trade agreements.

²⁰See https://en.wikipedia.org/wiki/Customs_union

²¹See <http://faculty.haas.berkeley.edu/arose/>

²² We have tested for the presence of the holdup problem in earliest period from 1950 to 1993 – available on demand – and we have found similar effects.

Notice that the last two terms are colinear to importer and exporter fixed effects respectively. This implies that we can use Feenstra (2002) approach to obtain unbiased estimates of the coefficients of interest by using time-varying importers and exporters fixed effects.²³ The fixed effects also control for any other country specific characteristics, such as productivity, status of being a landlocked or coastal country, quality of roads, infrastructure of road, etc.

Following the existing literature (e.g. Anderson and Van Wincoop (2003)), we model the transport costs τ_{ij} as being a log-linear function of the distance between countries i and j .

$$\tau_{ij} = d_{ij}^{\rho} \quad (4.16)$$

Where d_{ij} is the distance between countries i and j . To capture other iceberg costs, we include standard bilateral controls, described in appendix 4.6. The transit rent $\tilde{\varphi}_{mij}$ is defined as in section 4.2.1. We start by considering the same transit rent ($\tilde{\varphi}_{mij} = \varphi$). In section 4.5.3, we consider several transit rents depending on the existence of trade agreements. Plugging the functional forms of equations 4.16 into equation 4.15, and allowing for additional factors, the empirical model becomes

$$\ln X_{ij} = a_1 \tilde{\mathbf{1}}_{mij} + a_2 \ln d_{ij} + b_j \delta_j + b_i \delta_i + \sum_k b_k \mathbf{1}_{\alpha_k=1} + \varepsilon_{ij} \quad (4.17)$$

Where α_k is a set of bilateral dummy variables corresponding to additional controls such as trade agreements or common currency. δ_i is an indicator variable that is unity if country i exports, δ_j is an indicator variable that is unity if country j imports, and $a_1 = (1 - \sigma) \ln \varphi$. Given an elasticity of substitution σ , one can find the tariff-equivalent cost generated by the transit rents as

$$\varphi = e^{\frac{a_1}{1-\sigma}} \quad (4.18)$$

This regression approach allows us to show the effect of the transit rents on trade and to document the economic significance of the holdup problem.

²³With approximately 200 countries over 25 years, that amounts to over 5,000 fixed effects.

It stays silent, however, on any welfare implications. To understand further the distortionary effects of the holdup problem, one needs to study the general equilibrium consequences of transit rents on global welfare through their effect of prices. To do so, we proceed in two steps. The first step aims at calibrating the model's parameters in order to subsequently perform welfare analysis. Calibrating the model means finding the unobservable production Y_i (GDP excluding transit rents) of each country and the set of prices. We first estimate the transportation costs and transit rents using the regression results and the global network of trade routes. From the regression's coefficients, we back out the parameters φ and τ_{ij} using the functional forms described previously. With the information provided by the global network of trade route, one can then estimate the transit rents extracted by the most likely transit countries, which in turns make it possible to find the nominal productions $p_i Y_i$ using equation 4.8. We then solve for the prices $(\beta_i p_i)^{1-\sigma}$ and price indices P_i by solving the system of equations 4.9, 4.10 and 4.12 to find the set of prices. We finally recover the production Y_i of each country using prices and nominal productions. The nominal production is the fixed point of our subsequent welfare analyzes.

The first analysis we perform is to find prices and wealth in a world where the holdup friction and transit rents are not present. Doing so requires solving the same system if equations 4.9, 4.10 and 4.12, with fixed productions. This allows the determination of prices and wealth that are consistent with the production and trade frictions as estimated in the calibration. The second exercise is an experiment in which we compare the impact of transit rent as opposed to traditional transportation – ‘iceberg’ – costs.

4.5 Results

4.5.1 Existence of transit rents

Table 4.3 presents evidence of the existence and economic significance of the hold-up problem on trade. We use the 1993-2016 period as it exhibits substantially more data for countries that are most likely to be held up,

such as landlocked countries. The results hold for each year of the sample as shown in tables 4.11, 4.12 and 4.13 in the appendix, where we perform the same analysis on a yearly frequency. Every test include time, country, and time-varying fixed effect to capture the effect of the multilateral price pressures (see Feenstra (2002)) and all standard errors are clustered at the country pair level.

	Log(Trade)				
	Great-Circle Distance		Network Distance		
	(1)	(2)	(3)	(4)	(5)
Holdup		-0.940*** (0.062)	-0.825*** (0.063)	-0.807*** (0.068)	-0.836*** (0.069)
Log(Distance)	-1.581*** (0.022)	-1.550*** (0.021)			
Log(Distance)			-1.414*** (0.022)	-1.352*** (0.023)	-1.463*** (0.025)
Land (%)			-0.154 (0.179)	-0.157 (0.180)	-0.151 (0.198)
FTA	0.686*** (0.040)	0.644*** (0.039)	0.733*** (0.041)	0.710*** (0.049)	0.730*** (0.044)
WTO	0.852*** (0.067)	0.816*** (0.065)	0.758*** (0.067)	0.648*** (0.068)	0.856*** (0.095)
Timespan	1993–2016	1993–2016	1993–2016	1993–2004	2005–2016
Time FE	Y	Y	Y	Y	Y
Countries FE	Y	Y	Y	Y	Y
Time varying FE	Y	Y	Y	Y	Y
Observations	522,066	522,066	522,066	222,602	299,464
Adjusted R ²	0.725	0.726	0.722	0.716	0.726

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.3: Holdup Effect

The first two columns of table 4.3 use the great-circle distance while the last three make use of our new distance measure based on the global trade network. As described in the econometric model, the dependent variable

is the natural logarithm of the dollar amount of traded good between two partners. In all the paper's tests, we use exports, but as the data is symmetric, similar results can be seen when using imports instead. The 'hold up' independent variable is a dummy at one if the trade between the country pair is passing through a transit country. When using the network based distance, we are able to include control for the percentage of the trade route that is achieved on land as opposed to sea. All regressions include the control variables described in the appendix 4.6, but for the sake of clarity, only the coefficients regarding the presence of a bilateral trade agreement (foreign trade agreement or customs unions, variable 'FTA') and whether both countries trading belong to the World Trade Organization ('WTO') are reported. The first column serves as a control by excluding our variable of interest, the holdup dummy. This allows to observe that the coefficients of all the control variables are in line with the existing literature. For instance, the distance has a negative impact due to transportation costs, and bilateral agreements and being part of the WTO has a positive impact. We then add the holdup dummy (in column 2) and find that the presence of a transit country on the trade path indeed significantly affect trade. The third column makes use of our new distance measure to control more finely for realistic trade distances. It is important to note that the test performed in column 2 would not be possible without the construction of a global network of trade as the holdup dummy is created using that network. We observe that using more realistic distance estimations reduces the measured impact of the presence of transit countries. Indeed, when one uses the great-circle distance, the transportation costs are over-estimated because the great-circle distance underestimates long trading distances. Consequently, part of the effect of the holdup friction is wrongly captured by the distance measure. The overestimation of the transportation cost by the great-circle distance can be observed by looking at the difference between the second and the third column. Interestingly, it seems that the percentage of land on the trade route does not seem to significantly impact trade. That shows that an intuitive explanation as to the reason why landlocked countries trade less cannot simply be that they are far from the sea. The last two columns runs the same test as in column 3, but on

sub-samples. Column 4 focuses on the 1993-2004 subperiod and column 5 on 2005-2016. Both show that the holdup effect is consistent over time, a result generalized in our yearly estimation performed in the appendix. In terms of economic importance, the coefficient of the holdup dummy in column 3 (-0.825) tells us that on average, trade could be increased by approximately 28% for the hold up trades if they did not go through a transit country.²⁴

An important concern is that being held-up affects not only the amount of trade, but also the probability of initiating trade between two countries, resulting in a possible selection bias. To address this concern, we run a two stages Heckman procedure consisting of a first stage probit estimation using the same explanatory variables as the initial gravity specification. The second stage omits the common language dummy, shown by Helpman, M. Melitz, and Rubinstein (2008) to satisfy the exclusion restrictions. Due to computational costs, the first stage probit does not include time-varying country fixed effects, but only time and country fixed effects. Table 4.4 presents the same tests as table 4.3 with the Heckman procedure, using our network based distance measure only. The first two columns are respectively the first and second stages on the full 1993-2016 sample, and the last two columns provide the subperiods results. We find that while the selection bias is an issue as shown by the significant coefficient on the inverse Mill's ratio, our result are not significantly modified by correcting for the presence of zero-trades.

4.5.2 Global welfare implications

The above regression approach provides evidence of the existence and economic significance of the hold-up problem. However it does not help understand the welfare implication of the existence of transit rents. The global trade model developed in earlier sections enables us (i) to interpret more accurately the results of the previous regressions and (ii) to perform welfare analyses.

²⁴The percentage increase in trade if the holdup dummy goes from 1 to 0 is approximately $e^{(0.825)} - 1$.

	Does Trade (1) / Log(Trade) (2,3,4)			
	Full Sample		Sub-periods	
	(1)	(2)	(3)	(4)
Holdup (HU)	-0.160*** (0.009)	-0.866*** (0.064)	-0.842*** (0.069)	-0.877*** (0.070)
Log(Distance)	-0.495*** (0.003)	-1.507*** (0.026)	-1.427*** (0.026)	-1.569*** (0.028)
Land (%)	0.044* (0.026)	-0.122 (0.179)	-0.142 (0.181)	-0.106 (0.197)
FTA	0.396*** (0.009)	0.802*** (0.042)	0.758*** (0.051)	0.810*** (0.045)
WTO	0.149*** (0.006)	0.693*** (0.067)	0.592*** (0.069)	0.796*** (0.095)
Inverse Mill's ratio		0.476*** (0.058)	0.357*** (0.056)	0.547*** (0.066)
Timespan	1993–2016	1993–2016	1993–2004	2005–2016
Time FE	Y	Y	Y	Y
Countries FE	Y	Y	Y	Y
Time varying FE	N	Y	Y	Y
Observations	1,434,720	522,066	222,602	299,464
Adjusted R ²		0.721	0.715	0.724

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.4: Holdup Effect - Heckman Correction

As explained earlier, the variable $\tilde{\varphi}_{mij}$ in the model represents the transit fees and $\varphi - 1$ is the average tariff-equivalent cost of a trade being held up. This tariff-equivalent barrier cost can easily be retrieved from the previous regression coefficient using equation 4.18. However one needs to know the elasticity of substitution σ in order to find it. Section 4.2.4 provide a useful relationship between the elasticity of substitution and the transit fees that would be strategically changed by transit countries were they in a monopolistic position to do so. Combining equations 4.18 and 4.14, one can find the elasticity of substitution that is consistent with previous results if one consider that transit countries charge monopolistic fees. The elasticity of substitution considering our previous results is approximately $\sigma \approx 5$. This is in line with the elasticity of substitution considered in earlier studies such as in Anderson and Van Wincoop (2003). With $\sigma = 5$, the tariff-equivalent cost of being holdup is approximately 23%, that is, on average, transit countries charge 23 cents for every dollar worth of goods passing through. More than providing a finer interpretation of our previous results, the global trade model developed above allows us to perform welfare analyses. By calibrating the model's parameters consistent with the observed trade patterns, wealth and geographical constraint, we are able to analyze the general equilibrium consequences of the transit rents, and how the welfare would be modified were they not present. These calibrations are performed at a yearly frequency, and to observe a relatively stable number of countries over time, we are constraint by the availability of trade and GDP data. Therefore, we perform our welfare analyzes on the 2004-2016 period.

We perform two separate exercises. The first one is to estimate the global loss in utility due to the presence of the hold-up friction. The second is to analyse how this friction differs from transportation costs. For each exercise, we group the countries depending on whether they are landlocked countries or transit countries²⁵ The 'other' countries are the ones that are supposedly not affected by the holdup problem. Nonetheless, through the general equilibrium on prices, they are indirectly affected. We use three different measures

²⁵As explained above transit countries as classified as such when they are more often transit countries than holdup countries.

to understand the impact of the transit rents. First, we compute the change in welfare between a world where the holdup problem is present and a world where it is not. As explained in section 4.2.3, we use the equivalent variation to capture a change in welfare. The equivalent variation represents the monetary (dollar) equivalent that a given consumer loses due to the presence of transit rents. We report that wealth measure as a percentage of observed wealth. Second, we directly compare the loss in wealth in between the two worlds, over the full period 2004-2016 and also the yearly average. Finally, we measure the change in trade amount by computing the percentage change difference in the import to GDP ratio between the two worlds. Table 4.5 presents the results of the impact of transit rents on global welfare.

	All countries	Landlocked	Transit	Others
% Utility change	-0.15	-7.26	0.72	-0.12
Total Loss	-4521.75	-6503.24	4748.08	-2766.59
Yearly Loss	-347.83	-500.25	365.24	-212.81
% change in $\frac{Import}{GDP}$	-0.41	-7.56	0.21	-0.28

Losses are expressed in 2017 Billion USD

Table 4.5: General equilibrium effect of the holdup problem.

We find that the presence of transit rent reduces global welfare by approximately 0.15%. While this estimate might seem relatively small, the total loss over the 2004-2016 period amount to more than USD 4,500 billion, approximately USD 350 billion per year. More interestingly is the distribution of the welfare change among country groups. As expected, landlocked countries suffer the most by being heldup the most often, and the transit countries gain from their ability to extract transit rents. That represents a fundamental difference between the holdup costs and the transportation costs as will be seen shortly. Interestingly, the loss suffered by the landlocked countries is greater than the gain of transit countries. Due to the general equilibrium impact of prices, other countries, that might at first sight appear indifferent to the existence of this problem, are indeed affected. We estimate that these other countries, such as the United States or the United Kingdoms, suffer a welfare loss of approximately 0.12%. The welfare loss over the all sample for these

countries amount to more than USD 2,500 billion (a little more than USD 200 billion per year), which is a non-negligible impact. The main take-away of this exercise is that countries that are not directly held up or transit countries are still affected by this friction through the general equilibrium impact on prices.²⁶

The next natural question is to see how this problem compares to well known trade frictions such as transportation costs. To do so, we analyze how the different groups of countries would be affected if transportation cost increased in a way that would result in the same welfare loss. Specifically, we remove any hold-up friction and transit rents, and we adjust the model's transportation costs such that the global welfare loss due to increased transportation costs is equal to the welfare loss resulting from the presence of transit rents. The results of that exercise are shown in table 4.6.

	All countries	Landlocked	Transit	Others
% Utility change	-0.15	-0.49	-0.22	-0.12
Total Loss	-4521.75	-441.15	-1429.62	-2650.98
Yearly Loss	-347.83	-33.93	-109.97	-203.92
% change in $\frac{Import}{GDP}$	-0.10	-7.10	0.62	-0.02

Losses are expressed in 2017 Billion USD

¹ Transportation costs are adjusted such that a world without the holdup friction would generate an equivalent change in utility.

Table 4.6: General equilibrium effect for an equivalent change¹ in transportation costs.

The main difference is the distribution of gains and losses across the groups of countries. Unlike transit rents which represent a wealth transfer between holdup and transit countries, transportation costs are iceberg costs and are a wealth loss for all countries. Therefore an increase in transportation costs negatively impacts welfare for all country groups. As our estimate

²⁶Note that by featuring exogenous production, our model cannot capture the endogenous impact that lower transit rent may have on the production decision of held up countries. However, if anything, the welfare implications would be stronger if one allows such endogenous behavior, as the presence of transit rents most likely lowers incentives for higher production.

show landlocked countries are still more affected than transit countries because they are facing on average longer trading distances. But because there are no wealth transfer, both landlocked and transit countries are affected by transportation in the same magnitude. The impact of an increase in transportation costs for the other countries is particularly striking, as they seem to be affected as much, if not less, than by the presence of transit rents. This result speaks to the relevance of transit rents for these countries. While they are affected by transportation costs, and might therefore be incentivized to invest in infrastructure and technology to reduce these, our analyses shows that the existence of transit rents is as important for these countries. However, as they might not internalize the indirect impact that the hold-up problem generates on prices and their welfare, the incitation to mitigate this problem might be reduced.

4.5.3 Role of FTAs and customs unions

One concern about the previous exercise is that the transit rents have been globally estimated to be equal for every trade that are held up. However, the ability of the transit countries to extract transit rents might differ depending on the origin and destination of the goods transiting. For instance, one may rightly expect trade agreements between held up and transit countries to affect the rent extraction abilities of the transit countries. This section analyzes the impact of the two main agreement, trade agreements and customs unions.

We construct dummy variables to indicate whether there exists a free trade agreement between the holdup and the transit country for every held up trades for each year – as agreements are time varying – , or if both countries belong to the same customs union. When there are several transit countries on the trade path, we set the dummies at one if at least one of the transit country on the path has an active agreement with the holdup country. The European Union is of particular interest because it is features both a multilateral free trade agreement along with a customs union. Furthermore, the European Union allows free movement of good, capital and citizens within

the block, and one might wonder if it has the ability to mitigate the holdup problem above and beyond FTA and customs unions. We add a European Union dummy set at one if both holdup and transit countries are part of the E.U. and investigate if the E.U. provides any benefits in mitigating the problem. In order to have enough data and to avoid problems related to the expansion of the E.U. block, we run the tests using the 2004–2016 time period. Table 4.7 presents the results.

Columns 3 and 4 clearly illustrate the differential effect of FTAs and custom unions on the hold-up problem. It shows that the existence of a trade agreement between the holdup and the transit country does not seem to mitigate the problem. However, if both countries belong to the same customs union, the problem seems to clearly be reduced. This is in line with the fact that one way for transit countries to extract rent is to charge higher tariffs for holdup countries. However, if both countries belong to the same customs union, this possibility is reduced if not removed. The fact that customs union does not seem to completely eliminate the problem illustrate that transit countries may have other way to extract rent, such as highway tolls or others. In fact this result not only shows that customs union mitigate the problem, but it also reinforce the belief that transit countries engage in a rent extraction activity. Column 4 shows, unsurprisingly perhaps, that the problem is also reduced when the hold up and transit countries are part of the European Union. As column 5 and 6 show, and in line with the previous result, it seems that the positive impact of the European Union on the holdup problem is due to the existence of the European customs union and not from the free trade agreement. In addition it does not seem that the European Union is better able to mitigate the holdup problem than the presence of customs union. This suggests that for this particular problem, the European Union does not provide superior cooperation mechanisms than the one provided by the customs unions.

We then perform the same welfare analysis as before while taking into account the mitigating effect of customs unions. The impact of transit rents is shown in table 4.8 and the results with an increase in transportation costs are shown in table 4.9

	Log(Trade)					
	Base	FTA	CU	European Union		
	(1)	(2)	(3)	(4)	(5)	(6)
Holdup (HU)	-0.840*** (0.068)	-0.885*** (0.077)	-0.974*** (0.074)	-0.882*** (0.068)	-0.977*** (0.074)	-0.915*** (0.077)
HU X FTA		0.083 (0.083)				0.064 (0.082)
HU X CU			0.392*** (0.088)		0.344*** (0.097)	
HU X E.U.				0.311*** (0.117)	0.145 (0.127)	0.303*** (0.116)
Log(Distance)	-1.458*** (0.025)	-1.458*** (0.025)	-1.457*** (0.025)	-1.458*** (0.025)	-1.458*** (0.025)	-1.458*** (0.025)
Land (%)	-0.157 (0.197)	-0.165 (0.197)	-0.240 (0.196)	-0.186 (0.197)	-0.243 (0.196)	-0.192 (0.197)
FTA	0.740*** (0.044)	0.740*** (0.044)	0.739*** (0.044)	0.738*** (0.044)	0.738*** (0.044)	0.738*** (0.044)
WTO	0.840*** (0.093)	0.840*** (0.093)	0.841*** (0.093)	0.844*** (0.093)	0.843*** (0.093)	0.844*** (0.093)
Timespan	2004–2016	2004–2016	2004–2016	2004–2016	2004–2016	2004–2016
Time FE	Y	Y	Y	Y	Y	Y
Countries FE	Y	Y	Y	Y	Y	Y
Time varying FE	Y	Y	Y	Y	Y	Y
Observations	321,720	321,720	321,720	321,720	321,720	321,720
Adjusted R ²	0.726	0.726	0.726	0.726	0.726	0.726

Note:

*p<0.1; **p<0.05; ***p<0.01

'HU X FTA' ('HU X CU') is a dummy variable equal to 1 if a Free Trade Agreement (Customs Union) is active between the heldup and the transit country. 'HU X E.U.' is a dummy equal to 1 if both heldup and transit countries belong to the European Union.

Switzerland, Norway, Iceland and Liechtenstein are considered part of the E.U. in this analysis.

Table 4.7: Holdup Effect - Regional Trade Agreements

	All countries	Landlocked	Transit	Others
% Utility change	-0.12	-6.28	0.66	-0.11
Total Loss	-3660.05	-5643.21	4396.43	-2413.27
Yearly Loss	-281.54	-434.09	338.19	-185.64
% change in $\frac{Import}{GDP}$	-0.36	-6.63	0.17	-0.25

Losses are expressed in 2017 Billion USD

Table 4.8: General equilibrium effect of the holdup problem accounting for Customs Unions.

	All countries	Landlocked	Transit	Others
% Utility change	-0.12	-0.40	-0.17	-0.10
Total Loss	-3660.05	-360.68	-1157.60	-2141.77
Yearly Loss	-281.54	-27.74	-89.05	-164.75
% change in $\frac{Import}{GDP}$	-0.11	-6.27	0.50	-0.04

Losses are expressed in 2017 Billion USD

¹ Transportation costs are adjusted such that a world without the holdup friction would generate an equivalent change in utility.

Table 4.9: General equilibrium effect accounting for Customs Unions, for an equivalent change¹ in transportation costs.

The mitigating effect of customs unions does not modify the magnitude of the results albeit exhibiting smaller welfare losses. Interestingly, while the existence of customs unions reduces the average yearly loss by USD 70 billion (from 350 to 280) the loss for countries that are neither landlocked or transit countries is reduced only by less than USD 15 billion (from 200 to 185). Consistent with the previous welfare results, these other countries are more affected by the holdup problem than by an increase in transportation costs that would result in an equivalent global welfare loss.

4.5.4 Further discussion and robustness

In this section, we perform two additional tests for robustness purposes. The first one is to consider the possibility that both trade partners are being hold up by different transit countries. The hold up dummy used in the regression analyses is set at one when at least one of the trading partner is

being held up. However, it is possible that both partners are independently begin held up by two different transit countries. We show here that this possibility does not seem to affect the main result. The second test looks at the robustness of the holdup problem for different regions of the world. This robustness test is necessary to see if the global holdup problem identified in the previous sections is driven by a particular region. Table 4.10 presents these two robustness tests.

Column 2 shows that the fact that the trade path is double hold up does not affect the results. It also seems to suggest that trades that are double heldup do not seem to significantly suffer more than trade that are heldup by only one transit country. Column 3 provides evidence that the hold-up problem seems to be present and significant in all four continent in the world where there are hold up countries. However, it shows that the problem is more acute in Africa than in Europe for instance. One of the reason why the problem is mitigated in Europe is most likely due to the existence of the European Union and the presence of the European customs unions that, as has been shown above, mitigates the problem.

The appendix provides further robustness checks. Tables 4.11,4.12 and 4.13 provide yearly regressions of the main test on the existence of the holdup problem, and exhibit very stable coefficients of the holdup dummy over time. Tables 4.14, 4.15 and 4.16 are the same tests including the Heckman two stage procedure. Again, the results are stable and consistent with previous estimates. Finally, table 4.17 provide the robutness test presented in this section with the Heckman procedure. The result are in all accounts similar to the ones in table 4.10.

4.6 Conclusion

This paper provides a theoretical and empirical analysis of the trading costs created by the presence of transit countries in trading relationships. Because transit countries are in a unique position to extract rent from goods passing through, it negatively affects trades from and to holdup countries. Our theoretical analysis shows that such costs are fundamentally different from

	Log(Trade)		
	(1)	(2)	(3)
Holdup (HU)	-0.825*** (0.063)	-0.839*** (0.071)	
Double Holdup		-0.038 (0.074)	
HU Africa			-0.976*** (0.137)
HU Asia			-0.656*** (0.143)
HU Europe			-0.345*** (0.114)
HU America			-1.167*** (0.260)
Log(Sea+Land Distance)	-1.414*** (0.022)	-1.415*** (0.023)	-1.413*** (0.023)
Share of Land in distance	-0.154 (0.179)	-0.163 (0.180)	-0.221 (0.188)
FTA	0.733*** (0.041)	0.732*** (0.041)	0.725*** (0.041)
WTO	0.758*** (0.067)	0.759*** (0.067)	0.770*** (0.068)
Timespan	1993–2016	1993–2016	1993–2016
Time FE	Y	Y	Y
Countries FE	Y	Y	Y
Time varying FE	Y	Y	Y
Observations	522,066	522,066	522,066
Adjusted R ²	0.722	0.722	0.721

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.10: Holdup Effect - Robustness

transportation costs as they represent a wealth transfer between holdup and transit countries. Through its general equilibrium effect on prices, countries that are not directly held up are still affected by this holdup problem and bear a large share of the cost of the distortions created by this friction. By creating a global network of trade routes, we are able to accurately estimate trade distances and identify the most likely transit countries. This allows us to estimate that on average, these countries suffer from a yearly loss of more than USD 200 billion. Our empirical analyses show that the hold-up problem is partly mitigated when the held up and transit countries belong to the same customs union as it limits the ability of the transit countries to apply differential tariffs, arguably one of the main ways for them to extract rents. We do not find evidence that free trade agreements exhibit the same mitigating effect. While helpful, the presence of a customs union does not seem to drastically reduce the global welfare implications of the friction. We also provide evidence that the holdup friction is no less important for countries that are not directly held up than is the presence of transportation costs. While this paper stays relatively silent on the ways in which transit countries can extract rents, apart from the use of tariffs as our results on customs unions suggest, further research is needed to understand this complex problem in depth. While transport costs can be reduced through the use of infrastructure development and technology, the hold-up problem is more difficult to mitigate as it represents a complex international and political economy problem.

Bibliography

- Acharya, Viral V. and Lasse Heje Pedersen (2005). “Asset Pricing with Liquidity Risk”. *Journal of Financial Economics* 77.2, pp. 375–410.
- Acharya, Viral V. and Philipp Schnabl (May 1, 2009). “How Banks Played the Leverage “Game””. *Financial Markets, Institutions & Instruments* 18.2, pp. 144–145.
- Agarwal, Sumit and Itzhak Ben-David (Feb. 2014). *Loan Prospecting and the Loss of Soft Information*. Working Paper 19945. National Bureau of Economic Research.
- Allen, Franklin (Mar. 1, 1990). “The Market for Information and the Origin of Financial Intermediation”. *Journal of Financial Intermediation* 1.1, pp. 3–30.
- Allen, Franklin and Douglas Gale (1994). “Limited Stock Market Participation and Volatility of Asset Prices”. *American Economic Review* 84.4, pp. 933–955. JSTOR: 2118039.
- (2007). *Understanding Financial Crises*. Oxford University Press.
- Anderson, James E. and Eric Van Wincoop (2003). “Gravity with Gravititas: A Solution to the Border Puzzle”. *American Economic Review* 93.1, pp. 170–192.
- Arkolakis, Costas (2010). “Market Penetration Costs and the New Consumers Margin in International Trade”. *Journal of Political Economy* 118.6, pp. 1151–1199.
- Arkolakis, Costas et al. (May 2008). “Endogenous Variety and the Gains from Trade”. *American Economic Review* 98.2, pp. 444–450.

- Armington, Paul S. (1969). “A Theory of Demand for Products Distinguished by Place of Production”. *Staff Papers (International Monetary Fund)* 16.1, pp. 159–178.
- Arvis, Jean Francois, Gael Raballand, and Jean-Francois Marteau (2007). *The Cost of Being Landlocked: Logistics Costs And Supply Chain Reliability*. Working Paper 4258, pp. 1–81.
- Atkin, David and Dave Donaldson (July 2015). *Who’s Getting Globalized? The Size and Implications of Intra-National Trade Costs*. Working Paper 21439.
- Basel Committee (2013). “Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools”. *Basel Committee on Banking Supervision* (January), pp. 1–75.
- Bennardo, Alberto, Marco Pagano, and Salvatore Piccolo (Mar. 1, 2015). “Multiple Bank Lending, Creditor Rights, and Information Sharing”. *Review of Finance* 19.2, pp. 519–570.
- Berg, Tobias, Valentin Burg, et al. (Apr. 2018). *On the Rise of FinTechs - Credit Scoring Using Digital Footprints*. Working Paper 24551. National Bureau of Economic Research.
- Berg, Tobias, Manju Puri, and Jörg Rocholl (May 2013). *Loan Officer Incentives and the Limits of Hard Information*. Working Paper 19051. National Bureau of Economic Research.
- (Oct. 25, 2016). *Loan Officer Incentives, Internal Rating Models and Default Rates*. Working Paper ID 2022972. Rochester, NY: Social Science Research Network.
- Berger, Allen N., W. Scott Frame, and Nathan H. Miller (2005). “Credit Scoring and the Availability, Price, and Risk of Small Business Credit”. *Journal of Money, Credit and Banking* 37.2, pp. 191–222. JSTOR: 3838924.
- Bernard, Andrew B. et al. (Sept. 2003). “Plants and Productivity in International Trade”. *American Economic Review* 93.4, pp. 1268–1290.
- Bolton, Patrick and David S. Scharfstein (Mar. 1, 1990). “A Theory of Predation Based on Agency Problems in Financial Contracting”. *The American Economic Review* 80.1. ArticleType: research-article / Full publication

- date: Mar., 1990 / Copyright © 1990 American Economic Association, pp. 93–106. JSTOR: 2006736.
- Boot, Arnoud W. A. (Jan. 1, 2000). “Relationship Banking: What Do We Know?” *Journal of Financial Intermediation* 9.1, pp. 7–25.
- Brown, Martin, Tullio Jappelli, and Marco Pagano (Apr. 1, 2009). “Information Sharing and Credit: Firm-Level Evidence from Transition Countries”. *Journal of Financial Intermediation* 18.2, pp. 151–172.
- Brunnermeier, Markus K. and Lasse Heje Pedersen (2009). “Market Liquidity and Funding Liquidity”. *Review of Financial Studies* 22.6, pp. 2201–2238.
- Cabrales, Antonio, Olivier Gossner, and Roberto Serrano (2013). “Entropy and the Value of Information for Investors”. *American Economic Review* 103.1, pp. 360–377.
- Calomiris, Charles W., Florian Heider, and Marie Hoerova (2015). “A Theory of Bank Liquidity Requirement”.
- Cerasi, Vittoria and Sonja Daltung (Oct. 2000). “The Optimal Size of a Bank: Costs and Benefits of Diversification”. *European Economic Review* 44.9, pp. 1701–1726.
- Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (Aug. 1, 2017). *Does Improved Information Improve Incentives?* SSRN Scholarly Paper ID 2269380. Rochester, NY: Social Science Research Network.
- Chaney, Thomas (Sept. 2008). “Distorted Gravity: The Intensive and Extensive Margins of International Trade”. *American Economic Review* 98.4, pp. 1707–1721.
- Chemla, Gilles and Christopher A. Hennessy (Aug. 1, 2014). “Skin in the Game and Moral Hazard”. *The Journal of Finance* 69.4, pp. 1597–1641.
- Chemla, Gilles and Mario Milone (2018). “Oliver Hart - La Finance Vue à Travers La Théorie Des Contrats Incomplets”. *Les Grands Auteurs En Finance, 2ème Édition*. Éditions EMS.
- Coase, R.H. (1937). “The Nature of the Firm”. *Economica* 4.16, pp. 386–405.
- Costinot, Arnaud and Rodríguez-Clare Andrés (2014). “Trade Theory with Numbers: Quantifying the Consequences of Globalization”. *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Elsevier, pp. 197–261.

- Dewatripont, M. and E. Maskin (1995). “Credit and Efficiency in Centralized and Decentralized Economies”. *The Review of Economic Studies* 62.4, pp. 541–555.
- Diamond, Douglas (1984). “Financial Intermediation and Delegated Monitoring”. *The Review of Economic Studies* 51.3, pp. 393–414. JSTOR: 2297430.
- (1991). “Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt”. *Journal of Political Economy* 99.4, pp. 689–721.
- Diamond, Douglas W. and Philip H. Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity”. *Journal of Political Economy* 91.3, p. 401.
- Diamond, Douglas W. and Raghuram G. Rajan (2001). “Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking”. *Journal of Political Economy* 109, pp. 287–327.
- (2011). “Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes”. *Quarterly Journal of Economics* 126.2, pp. 557–591.
- Djankov, Simeon, Caroline Freund, and Cong S Pham (Jan. 8, 2010). “Trading on Time”. *The Review of Economics and Statistics* 92.1, pp. 166–173.
- Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer (May 1, 2007). “Private Credit in 129 Countries”. *Journal of Financial Economics* 84.2, pp. 299–329.
- Eaton, Jonathan and Samuel Kortum (2002). “Technology, Geography, and Trade”. *Econometrica* 70.5, pp. 1741–1779.
- Ellis, Andrew (Jan. 1, 2018). “Foundations for Optimal Inattention”. *Journal of Economic Theory* 173, pp. 56–94.
- Feenstra, Robert C. (2002). “Border Effects and the Gravity Equation: Consistent Methods for Estimation”. *Scottish Journal of Political Economy* 49.5, pp. 491–506.
- (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press.

- Feinberg, Yossi and Morton I. Kamien (2001). “Highway Robbery: Complementary Monopoly and the Hold-up Problem”. *International Journal of Industrial Organization* 19.10, pp. 1603–1621.
- Franck, Irene M. and David M. Brownstone (1986). *The Silk Road: A History*. Facts On File Inc. 356 pp.
- Frankel, Jeffrey A. and David Romer (1999). “Does Trade Cause Growth?” *American Economic Review* 89.3, pp. 379–399.
- Freixas, Xavier and Jean-Charles Rochet (2008). *Microeconomics of Banking*. The MIT Press.
- Froot, Kenneth A., David S Scharfstein, and Jeremy C. Stein (1993). “Risk Management: Coordinating Corporate Investment and Financing Policies”. *The Journal of Finance* 48, pp. 1629–1658.
- Froot, Kenneth A. and Jeremy C. Stein (1998). “Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach”. *Journal of Financial Economics* 47.1, pp. 55–82.
- Fuster, Andreas et al. (July 26, 2018). *Predictably Unequal? The Effects of Machine Learning on Credit Markets*. SSRN Scholarly Paper ID 3072038. Rochester, NY: Social Science Research Network.
- Gale, Douglas and Martin Hellwig (1985). “Incentive-Compatible Debt Contracts: The One-Period Problem”. *The Review of Economic Studies* 52.4, pp. 647–663. JSTOR: 2297737.
- Gardner, Roy et al. (2002). *Tolling the Rhine in 1254: Complementary Monopoly Revisited*. Working Paper.
- Gorton, Gary (2010). *Slapped by the Invisible Hand: The Panic of 2007*.
- Gorton, Gary and George Pennacchi (1990). “Financial Intermediaries and Liquidity Creation”. *The Journal of Finance* 45.1, pp. 49–71.
- (June 1, 1995). “Banks and Loan Sales Marketing Nonmarketable Assets”. *Journal of Monetary Economics* 35, pp. 389–411.
- Grossman, Sanford and Oliver Hart (1986). “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”. *Journal of Political Economy* 94.4, pp. 691–691.
- Hanson, Samuel et al. (2015). “Banks as Patient Fixed Income Investors”. *Journal of Financial Economics*.

- Hart, Oliver (1989). “An Economist’s Perspective on the Theory of the Firm”. *Columbia Law Review*.
- (2011). “Thinking about the Firm: A Review of Daniel Spulber’s The Theory of the Firm”. *Journal of Economic Literature* 49.1, pp. 101–113.
- Hart, Oliver D. and Dwight Jaffee (1974). “On the Application of Portfolio Theory to Depository Financial Intermediaries”. *The Review of Economic Studies* 41.1, pp. 129–147. JSTOR: 2296404.
- Hart, Oliver and John Moore (1990). “Property Rights and the Nature of the Firm”. *Journal of Political Economy* 98.6, pp. 1119–1158.
- Hartman-Glaser, Barney, Tomasz Piskorski, and Alexei Tchisty (2012). “Optimal Securitization with Moral Hazard”. *Journal of Financial Economics* 104.1, pp. 186–202.
- Head, Keith and Thierry Mayer (June 2010). “Illusory Border Effects”. *The Gravity Model in International Trade: Advances and Applications*. Ed. by Steven Brakamn and Peter van Bergeijk. Cambridge University Press.
- (Jan. 1, 2014). “Gravity Equations: Workhorse, Toolkit, and Cookbook”. *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Handbook of International Economics. Elsevier, pp. 131–195.
- Helpman, Elhanan, Marc Melitz, and Yona Rubinstein (2008). “Estimating Trade Flows: Trading Partners and Trading Volumes”. *The Quarterly Journal of Economics* 123.2, pp. 441–487.
- Hertzberg, Andrew, José María Liberti, and Daniel Paravisini (Apr. 1, 2011). “Public Information and Coordination: Evidence from a Credit Registry Expansion”. *The Journal of Finance* 66.2, pp. 379–412.
- Hirschman, Albert O. (1945). *National Power and the Structure of Foreign Trade*. University of California Press.
- Holmström, Bengt and Jean Tirole (2011). *Inside and Outside Liquidity*. MIT Press.
- Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba (2009). “The Trade Reducing Effects of Market Power in International Shipping”. *Journal of Development Economics* 89.1, pp. 84–97.

- Inderst, Roman and Holger M. Mueller (2006). “Informed Lending and Security Design”. *Journal of Finance* 61.5, pp. 2137–2162.
- (2007). “A Lender-Based Theory of Collateral”. *Journal of Financial Economics* 84.3, pp. 826–859.
- Innes, Robert D (Oct. 1, 1990). “Limited Liability and Incentive Contracting with Ex-Ante Action Choices”. *Journal of Economic Theory* 52.1, pp. 45–67.
- Jagtiani, Julapa and Catharine Lemieux (Apr. 5, 2018). *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the Lendingclub Consumer Platform*. SSRN Scholarly Paper ID 3178461. Rochester, NY: Social Science Research Network.
- Jappelli, Tullio and Marco Pagano (Oct. 1, 2002). “Information Sharing, Lending and Defaults: Cross-Country Evidence”. *Journal of Banking & Finance* 26.10, pp. 2017–2045.
- Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp (Mar. 1, 2016). “A Rational Theory of Mutual Funds’ Attention Allocation”. *Econometrica* 84.2, pp. 571–626.
- Karni, Edi and Subir K. Chakrabarti (1997). “Political Structure, Taxes and Trade”. *Journal of Public Economics* 64.2, pp. 241–258.
- Karolyi, Stephen Adam (Feb. 1, 2018). “Personal Lending Relationships”. *The Journal of Finance* 73.1, pp. 5–49.
- Kashyap, Anil K., Raghuram G. Rajan, and Jeremy C. Stein (2002). “Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking”. *The Journal of Finance* 57.1, pp. 33–73.
- Kim, Daesik and Anthony M Santomero (1988). “Risk in Banking and Capital Regulation”. *Journal of Finance* 43.5, pp. 1219–1233.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978). “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process”. *The Journal of Law & Economics* 21.2, pp. 297–326.
- Koehn, Michael and Anthony M Santomero (1980). “Regulation of Bank Capital and Portfolio Risk”. *Journal of Finance* 35, pp. 1235–1244.
- Krugman, Paul (1980). “Scale Economies, Product Differentiation, and the Pattern of Trade.” *American Economic Review* 70.5, pp. 950–959.

- Levitt, Steven D. and Christopher M. Snyder (1997). "Is No News Bad News? Information Transmission and the Role of "Early Warning" in the Principal-Agent Model". *The RAND Journal of Economics* 28.4, pp. 641–661. JSTOR: 2555780.
- Liberti, José María and Mitchell A. Petersen (Mar. 2017). *Information: Hard and Soft*. Working Paper.
- Limão, Nuno and Anthony J. Venables (2001). "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade". *The World Bank Economic Review* 15.3, pp. 451–479.
- Maggi, Giovanni (Jan. 1, 2014). "Chapter 6 - International Trade Agreements". *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Handbook of International Economics. Elsevier, pp. 317–390.
- Markowitz, Harry (1952). "Portfolio Selection". *The Journal of Finance* 7, pp. 77–91.
- Matejka, Filip and Alisdair McKay (Jan. 2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model". *American Economic Review* 105.1, pp. 272–298.
- Mayskaya, Tatiana (Aug. 4, 2017). *Dynamic Choice of Information Sources*. SSRN Scholarly Paper ID 2863605. Rochester, NY: Social Science Research Network.
- Melitz, Marc J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". *Econometrica* 71.6, pp. 1695–1725.
- Mester, Loretta J., Leonard I. Nakamura, and Micheline Renault (May 1, 2007). "Transactions Accounts and Loan Monitoring". *The Review of Financial Studies* 20.3, pp. 529–556.
- Meyer, Margaret A. (1991). "Learning from Coarse Information: Biased Contests and Career Profiles". *The Review of Economic Studies* 58.1, pp. 15–41. JSTOR: 2298043.
- Meyer, Margaret A. and Jeffrey Zwiebel (Mar. 2007). *Learning and Self-Reinforcing Behavior*. Working Paper.

- Miyagiwa, Kaz (Apr. 2009). *The Silk Road: Tax Competition among Nation States*. 18. Center for Research on Contemporary Economic Systems, Graduate School of Economics, Hitotsubashi University.
- Nimark, Kristoffer P. and Savitar Sundaresan (2018). *Inattention and Belief Polarization*. Working Paper.
- Norden, Lars and Martin Weber (Oct. 1, 2010). “Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers”. *The Review of Financial Studies* 23.10, pp. 3665–3699.
- Overman, Henry G., Stephen Redding, and Anthony J. Venables (2008). “The Economic Geography of Trade, Production, and Income: A Survey of Empirics”. *Handbook of International Trade*. Ed. by E. Kwan Choi and James Harrigan. Wiley-Blackwell, pp. 350–387.
- Padilla, A. Jorge and Marco Pagano (Dec. 1, 2000). “Sharing Default Information as a Borrower Discipline Device”. *European Economic Review* 44.10, pp. 1951–1980.
- Paravisini, Daniel and Antoinette Schoar (2015). *The Incentive Effect of Scores: Randomized Evidence from Credit Committees*. Working Paper 19303. National Bureau of Economic Research.
- Parlour, Christine A. and Guillaume Plantin (June 1, 2008). “Loan Sales and Relationship Banking”. *The Journal of Finance* 63.3, pp. 1291–1314.
- Pennacchi, George (2012). “Narrow Banking”. *Annual Review of Financial Economics* 4.0.
- Petersen, Mitchell A. and Raghuram Rajan (1994). “The Benefits of Lending Relationships: Evidence from Small Business Data”. *The Journal of Finance* 49.1, pp. 3–37.
- (May 1, 1995). “The Effect of Credit Market Competition on Lending Relationships”. *The Quarterly Journal of Economics* 110.2, pp. 407–443.
- (2002a). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *Journal of Finance* 57.6, pp. 2533–2570.
- (Dec. 1, 2002b). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *The Journal of Finance* 57.6, pp. 2533–2570.
- Philippon, Thomas (Aug. 2016). *The FinTech Opportunity*. Working Paper 22476. National Bureau of Economic Research.

- Puri, Manju, Jörg Rocholl, and Sascha Steffen (July 1, 2017). “What Do a Million Observations Have to Say about Loan Defaults? Opening the Black Box of Relationships”. *Journal of Financial Intermediation* 31, pp. 1–15.
- Pyle, David H. (1971). “On the Theory of Financial Intermediation”. *The Journal of Finance* 26.3, pp. 737–747.
- Raballand, Gael J. R. F. and Supee Teravaninthorn (Oct. 10, 2008). *Transport Prices and Costs in Africa : A Review of the International Corridors*. 46181. The World Bank, pp. 1–166.
- Raballand, Gaël, Antoine Kunth, and Richard Auty (Mar. 1, 2005). “Central Asia’s Transport Cost Burden and Its Impact on Trade”. *Economic Systems* 29.1, pp. 6–31.
- Raballand, Gael et al. (2012). “Why Cargo Dwell Time Matters in Trade”. *Economic Premise, World Bank* 81.
- Rajan, Raghuram (1992). “Insiders and Outsiders: The Choice between Informed and Arm’s-Length Debt”. *The Journal of Finance* 47.4, pp. 1367–1400.
- Rajan, Uday, Amit Seru, and Vikrant Vig (Feb. 1, 2015). “The Failure of Models That Predict Failure: Distance, Incentives, and Defaults”. *Journal of Financial Economics* 115.2, pp. 237–260.
- Redding, Stephen and Anthony J. Venables (Jan. 1, 2004). “Economic Geography and International Inequality”. *Journal of International Economics* 62.1, pp. 53–82.
- Repullo, Rafael (2005). “Liquidity , Risk Taking , and the Lender of Last Resort”. *International Journal of Central Banking* 1.2, pp. 47–80.
- Rochet, Jean-Charles (1992). “Capital Requirements and the Behaviour of Commercial Banks”. *European Economic Review* 36, pp. 1137–1170.
- Rose, Andrew K. (2004). “Do WTO Members Have More Liberal Trade Policy?” *Journal of International Economics* 63.2, pp. 209–235.
- Shleifer, Andrei and Robert W. Vishny (1992). “Liquidation Values and Debt Capacity: A Market Equilibrium Approach”. *Journal of Finance* 47.4, pp. 1343–1366.

- Shorrer, Ran I. (Dec. 2015). *Entropy and the Value of Informatio for Investors: The Prior-Free Implications*. Working Paper.
- Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Reprint, Liberty Classics, 1979.
- Stein, Jeremy C (2012). “Monetary Policy as Financial Stability Regulation”. *The Quarterly Journal of Economics* 127, pp. 57–95.
- Stein, Jeremy C. (2002). “Information Production and Capital Allocation: Decentralized versus Hierarchical Firms”. *The Journal of Finance* 57.5, pp. 1891–1921.
- Townsend, Robert M (Oct. 1, 1979). “Optimal Contracts and Competitive Markets with Costly State Verification”. *Journal of Economic Theory* 21.2, pp. 265–293.
- USAID (2004). *The Broad Economic Impact of Port Inefficiency: A Comparative Study of Two Ports*. Washington DC.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010). “Information Acquisition and Under-Diversification”. *The Review of Economic Studies* 77.2, pp. 779–805.
- Williamson, Oliver E. (1971). “The Vertical Integration of Production: Market Failure Considerations”. *The American Economic Review* 61.2, pp. 112–123.
- (1975). *Markets and Hierarchies: Analysis and Antitrust Implications : A Study in the Economics of Internal Organization*. Free Press. 312 pp.

Appendix

Data and control variables

All regressions include the following country-pair specific control variables.

- Common Language: Dummy at 1 if both countries use the same language.
- Common Colonizer: Dummy at 1 if both countries were or are colonized by the same colonizer.
- Colonies: Dummy at 1 if the exporter is a colony of the importer at time t or vice-versa.
- Ever Colonized: Dummy at 1 if the exporter or the importer were ever colonized.
- Common Currency: Dummy at 1 if both countries use the same currency at time t .
- FTA: Dummy at 1 if there is a bilateral trade agreement between both countries at time t or if both countries are part of a multilateral trade agreement at time t .
- WTO: Dummy at 1 if both countries belong to the World Trade Organization at time t .
- One WTO: Dummy at 1 if only one country belongs to the World Trade Organization.

- GSP: Dummy at 1 if the exporter belongs to the list of the importer's Generalized Schemes of Preferences at time t or vice-versa.
- Number of islands: Number of islands in the pair of countries (0, 1 or 2).

Additional Tables

Table 4.11: Holdup Effect per Year

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.901*** (0.106)	-0.733*** (0.100)	-0.817*** (0.101)	-0.965*** (0.099)	-0.819*** (0.090)	-0.789*** (0.087)	-0.752*** (0.088)	-0.789*** (0.089)
Log(Sea+Land Distance)	-1.289*** (0.032)	-1.314*** (0.029)	-1.345*** (0.029)	-1.350*** (0.029)	-1.343*** (0.029)	-1.312*** (0.027)	-1.330*** (0.027)	-1.367*** (0.029)
Share of Land in distance	0.285 (0.237)	-0.109 (0.218)	0.111 (0.221)	0.205 (0.220)	-0.099 (0.215)	-0.209 (0.224)	-0.448* (0.234)	-0.260 (0.217)
FTA	0.392*** (0.085)	0.519*** (0.078)	0.422*** (0.075)	0.454*** (0.072)	0.481*** (0.068)	0.590*** (0.064)	0.599*** (0.066)	0.802*** (0.065)
WTO	0.893*** (0.101)	0.821*** (0.104)	0.790*** (0.106)	0.749*** (0.107)	0.573*** (0.105)	0.454*** (0.105)	0.340*** (0.109)	0.321*** (0.112)
Timespan	1993	1994	1995	1996	1997	1998	1999	2000
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	14,332	15,214	15,877	16,501	17,454	17,973	18,740	20,500
Adjusted R ²	0.704	0.708	0.710	0.705	0.711	0.719	0.719	0.711

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.12: Holdup Effect per Year

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.695*** (0.091)	-0.739*** (0.086)	-0.809*** (0.089)	-0.878*** (0.088)	-0.770*** (0.090)	-0.804*** (0.089)	-0.751*** (0.088)	-0.793*** (0.090)
Log(Sea+Land Distance)	-1.398*** (0.030)	-1.393*** (0.029)	-1.365*** (0.029)	-1.379*** (0.030)	-1.378*** (0.029)	-1.403*** (0.029)	-1.401*** (0.030)	-1.422*** (0.030)
Share of Land in distance	-0.304 (0.225)	-0.301 (0.227)	-0.153 (0.232)	-0.245 (0.240)	-0.103 (0.217)	-0.119 (0.229)	-0.140 (0.236)	-0.055 (0.239)
FTA	0.820*** (0.065)	0.831*** (0.063)	0.903*** (0.064)	0.905*** (0.063)	0.873*** (0.063)	0.921*** (0.062)	0.940*** (0.062)	1.006*** (0.062)
WTO	0.625*** (0.129)	0.654*** (0.132)	0.751*** (0.131)	0.636*** (0.133)	0.976*** (0.129)	0.858*** (0.124)	0.838*** (0.128)	0.795*** (0.130)
Timespan	2001	2002	2003	2004	2005	2006	2007	2008
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	20,783	21,143	21,829	22,256	22,728	23,415	23,874	24,502
Adjusted R ²	0.716	0.715	0.722	0.726	0.726	0.728	0.726	0.724

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.13: Holdup Effect per Year

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.834*** (0.086)	-0.825*** (0.086)	-0.800*** (0.085)	-0.901*** (0.088)	-0.838*** (0.089)	-0.932*** (0.086)	-0.866*** (0.088)	-0.877*** (0.086)
Log(Sea+Land Distance)	-1.470*** (0.029)	-1.495*** (0.029)	-1.503*** (0.029)	-1.474*** (0.029)	-1.497*** (0.030)	-1.514*** (0.029)	-1.468*** (0.029)	-1.484*** (0.028)
Share of Land in distance	0.046 (0.230)	-0.124 (0.231)	-0.088 (0.229)	-0.080 (0.227)	-0.222 (0.229)	-0.059 (0.231)	-0.260 (0.234)	-0.571** (0.230)
FTA	0.607*** (0.058)	0.610*** (0.056)	0.626*** (0.056)	0.734*** (0.053)	0.718*** (0.053)	0.656*** (0.050)	0.689*** (0.050)	0.642*** (0.049)
WTO	0.838*** (0.122)	0.915*** (0.122)	0.740*** (0.127)	0.732*** (0.142)	0.851*** (0.146)	0.901*** (0.145)	0.813*** (0.154)	1.038*** (0.145)
Timespan	2009	2010	2011	2012	2013	2014	2015	2016
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	24,553	24,966	25,376	25,488	25,735	25,987	26,205	26,635
Adjusted R ²	0.728	0.730	0.726	0.726	0.724	0.726	0.722	0.725

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.14: Holdup Effect per Year - Heckman Correction

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.936*** (0.106)	-0.772*** (0.101)	-0.846*** (0.102)	-0.994*** (0.099)	-0.854*** (0.091)	-0.813*** (0.088)	-0.777*** (0.088)	-0.828*** (0.090)
Log(Sea+Land Distance)	-1.347*** (0.035)	-1.378*** (0.031)	-1.411*** (0.031)	-1.406*** (0.031)	-1.408*** (0.032)	-1.362*** (0.029)	-1.375*** (0.029)	-1.460*** (0.032)
Share of Land in distance	0.299 (0.236)	-0.098 (0.218)	0.110 (0.222)	0.210 (0.221)	-0.092 (0.216)	-0.206 (0.225)	-0.440* (0.234)	-0.247 (0.218)
FTA	0.394*** (0.088)	0.539*** (0.080)	0.451*** (0.077)	0.476*** (0.073)	0.491*** (0.069)	0.645*** (0.066)	0.666*** (0.067)	0.859*** (0.066)
WTO	0.891*** (0.102)	0.792*** (0.104)	0.772*** (0.106)	0.724*** (0.107)	0.554*** (0.106)	0.432*** (0.105)	0.325*** (0.109)	0.240** (0.112)
Inverse Mill's ratio	0.224*** (0.075)	0.255*** (0.073)	0.288*** (0.073)	0.231*** (0.073)	0.293*** (0.077)	0.249*** (0.077)	0.206*** (0.070)	0.477*** (0.075)
Timespan	1993	1994	1995	1996	1997	1998	1999	2000
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	14,332	15,214	15,877	16,501	17,454	17,973	18,740	20,500
Adjusted R ²	0.703	0.706	0.709	0.704	0.710	0.717	0.717	0.710

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.15: Holdup Effect per Year - Heckman Correction

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.735*** (0.092)	-0.792*** (0.087)	-0.842*** (0.090)	-0.901*** (0.089)	-0.794*** (0.091)	-0.827*** (0.090)	-0.768*** (0.089)	-0.825*** (0.090)
Log(Sea+Land Distance)	-1.495*** (0.032)	-1.488*** (0.032)	-1.477*** (0.032)	-1.475*** (0.032)	-1.471*** (0.032)	-1.499*** (0.032)	-1.498*** (0.032)	-1.530*** (0.033)
Share of Land in distance	-0.296 (0.226)	-0.298 (0.227)	-0.131 (0.232)	-0.205 (0.239)	-0.056 (0.217)	-0.058 (0.229)	-0.105 (0.235)	-0.030 (0.238)
FTA	0.877*** (0.066)	0.884*** (0.064)	1.030*** (0.065)	1.019*** (0.064)	0.988*** (0.064)	1.016*** (0.063)	1.042*** (0.063)	1.109*** (0.063)
WTO	0.530*** (0.128)	0.540*** (0.131)	0.682*** (0.131)	0.568*** (0.132)	0.914*** (0.128)	0.813*** (0.124)	0.793*** (0.127)	0.755*** (0.129)
Inverse Mill's ratio	0.538*** (0.075)	0.538*** (0.074)	0.587*** (0.079)	0.508*** (0.080)	0.467*** (0.078)	0.501*** (0.081)	0.501*** (0.081)	0.581*** (0.082)
Timespan	2001	2002	2003	2004	2005	2006	2007	2008
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	20,783	21,143	21,829	22,256	22,728	23,415	23,874	24,502
Adjusted R ²	0.715	0.715	0.721	0.725	0.725	0.727	0.725	0.723

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.16: Holdup Effect per Year - Heckman Correction

	Log(Trade)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Holdup (HU)	-0.878*** (0.087)	-0.867*** (0.087)	-0.840*** (0.085)	-0.944*** (0.089)	-0.894*** (0.090)	-0.988*** (0.087)	-0.927*** (0.090)	-0.942*** (0.087)
Log(Sea+Land Distance)	-1.590*** (0.032)	-1.615*** (0.032)	-1.608*** (0.032)	-1.588*** (0.031)	-1.624*** (0.032)	-1.620*** (0.032)	-1.594*** (0.031)	-1.599*** (0.030)
Share of Land in distance	0.091 (0.229)	-0.087 (0.229)	-0.065 (0.228)	-0.010 (0.227)	-0.161 (0.228)	-0.015 (0.232)	-0.242 (0.235)	-0.518** (0.230)
FTA	0.694*** (0.059)	0.684*** (0.056)	0.692*** (0.056)	0.810*** (0.054)	0.786*** (0.053)	0.731*** (0.051)	0.775*** (0.050)	0.711*** (0.050)
WTO	0.803*** (0.121)	0.883*** (0.122)	0.722*** (0.127)	0.689*** (0.142)	0.784*** (0.145)	0.829*** (0.145)	0.699*** (0.154)	0.929*** (0.144)
Inverse Mill's ratio	0.634*** (0.084)	0.629*** (0.083)	0.539*** (0.084)	0.597*** (0.086)	0.684*** (0.085)	0.590*** (0.084)	0.732*** (0.084)	0.672*** (0.083)
Timespan	2009	2010	2011	2012	2013	2014	2015	2016
Time FE	N	N	N	N	N	N	N	N
Countries FE	Y	Y	Y	Y	Y	Y	Y	Y
Time varying FE	N	N	N	N	N	N	N	N
Observations	24,553	24,966	25,376	25,488	25,735	25,987	26,205	26,635
Adjusted R ²	0.727	0.729	0.725	0.725	0.723	0.725	0.721	0.725

Note: *p<0.1; **p<0.05; ***p<0.01

	Log(Trade)		
	(1)	(2)	(3)
Holdup (HU)	-0.866*** (0.064)	-0.891*** (0.073)	
Double Holdup		-0.071 (0.075)	
HU Africa			-1.124*** (0.135)
HU Asia			-0.683*** (0.141)
HU Europe			-0.324*** (0.117)
HU America			-1.218*** (0.263)
Log(Sea+Land Distance)	-1.507*** (0.026)	-1.509*** (0.026)	-1.511*** (0.026)
Share of Land in distance	-0.122 (0.179)	-0.139 (0.180)	-0.188 (0.190)
FTA	0.802*** (0.042)	0.801*** (0.042)	0.791*** (0.042)
WTO	0.693*** (0.067)	0.694*** (0.067)	0.702*** (0.068)
Inverse Mill's ratio	0.476*** (0.058)	0.477*** (0.058)	0.514*** (0.058)
Timespan	1993–2016	1993–2016	1993–2016
Time FE	Y	Y	Y
Countries FE	Y	Y	Y
Time varying FE	Y	Y	Y
Observations	522,066	522,066	522,066
Adjusted R ²	0.721	0.721	0.720

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.17: Holdup Effect - Robustness - Heckman Correction

Bibliography

- Acharya, Viral V. and Lasse Heje Pedersen (2005). “Asset Pricing with Liquidity Risk”. *Journal of Financial Economics* 77.2, pp. 375–410.
- Acharya, Viral V. and Philipp Schnabl (May 1, 2009). “How Banks Played the Leverage “Game””. *Financial Markets, Institutions & Instruments* 18.2, pp. 144–145.
- Agarwal, Sumit and Itzhak Ben-David (Feb. 2014). *Loan Prospecting and the Loss of Soft Information*. Working Paper 19945. National Bureau of Economic Research.
- Allen, Franklin (Mar. 1, 1990). “The Market for Information and the Origin of Financial Intermediation”. *Journal of Financial Intermediation* 1.1, pp. 3–30.
- Allen, Franklin and Douglas Gale (1994). “Limited Stock Market Participation and Volatility of Asset Prices”. *American Economic Review* 84.4, pp. 933–955. JSTOR: 2118039.
- (2007). *Understanding Financial Crises*. Oxford University Press.
- Anderson, James E. and Eric Van Wincoop (2003). “Gravity with Gravititas: A Solution to the Border Puzzle”. *American Economic Review* 93.1, pp. 170–192.
- Arkolakis, Costas (2010). “Market Penetration Costs and the New Consumers Margin in International Trade”. *Journal of Political Economy* 118.6, pp. 1151–1199.
- Arkolakis, Costas et al. (May 2008). “Endogenous Variety and the Gains from Trade”. *American Economic Review* 98.2, pp. 444–450.

- Armington, Paul S. (1969). “A Theory of Demand for Products Distinguished by Place of Production”. *Staff Papers (International Monetary Fund)* 16.1, pp. 159–178.
- Arvis, Jean Francois, Gael Raballand, and Jean-Francois Marteau (2007). *The Cost of Being Landlocked: Logistics Costs And Supply Chain Reliability*. Working Paper 4258, pp. 1–81.
- Atkin, David and Dave Donaldson (July 2015). *Who’s Getting Globalized? The Size and Implications of Intra-National Trade Costs*. Working Paper 21439.
- Basel Committee (2013). “Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools”. *Basel Committee on Banking Supervision* (January), pp. 1–75.
- Bennardo, Alberto, Marco Pagano, and Salvatore Piccolo (Mar. 1, 2015). “Multiple Bank Lending, Creditor Rights, and Information Sharing”. *Review of Finance* 19.2, pp. 519–570.
- Berg, Tobias, Valentin Burg, et al. (Apr. 2018). *On the Rise of FinTechs - Credit Scoring Using Digital Footprints*. Working Paper 24551. National Bureau of Economic Research.
- Berg, Tobias, Manju Puri, and Jörg Rocholl (May 2013). *Loan Officer Incentives and the Limits of Hard Information*. Working Paper 19051. National Bureau of Economic Research.
- (Oct. 25, 2016). *Loan Officer Incentives, Internal Rating Models and Default Rates*. Working Paper ID 2022972. Rochester, NY: Social Science Research Network.
- Berger, Allen N., W. Scott Frame, and Nathan H. Miller (2005). “Credit Scoring and the Availability, Price, and Risk of Small Business Credit”. *Journal of Money, Credit and Banking* 37.2, pp. 191–222. JSTOR: 3838924.
- Bernard, Andrew B. et al. (Sept. 2003). “Plants and Productivity in International Trade”. *American Economic Review* 93.4, pp. 1268–1290.
- Bolton, Patrick and David S. Scharfstein (Mar. 1, 1990). “A Theory of Predation Based on Agency Problems in Financial Contracting”. *The American Economic Review* 80.1. ArticleType: research-article / Full publication

- date: Mar., 1990 / Copyright © 1990 American Economic Association, pp. 93–106. JSTOR: 2006736.
- Boot, Arnoud W. A. (Jan. 1, 2000). “Relationship Banking: What Do We Know?” *Journal of Financial Intermediation* 9.1, pp. 7–25.
- Brown, Martin, Tullio Jappelli, and Marco Pagano (Apr. 1, 2009). “Information Sharing and Credit: Firm-Level Evidence from Transition Countries”. *Journal of Financial Intermediation* 18.2, pp. 151–172.
- Brunnermeier, Markus K. and Lasse Heje Pedersen (2009). “Market Liquidity and Funding Liquidity”. *Review of Financial Studies* 22.6, pp. 2201–2238.
- Cabrales, Antonio, Olivier Gossner, and Roberto Serrano (2013). “Entropy and the Value of Information for Investors”. *American Economic Review* 103.1, pp. 360–377.
- Calomiris, Charles W., Florian Heider, and Marie Hoerova (2015). “A Theory of Bank Liquidity Requirement”.
- Cerasi, Vittoria and Sonja Daltung (Oct. 2000). “The Optimal Size of a Bank: Costs and Benefits of Diversification”. *European Economic Review* 44.9, pp. 1701–1726.
- Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (Aug. 1, 2017). *Does Improved Information Improve Incentives?* SSRN Scholarly Paper ID 2269380. Rochester, NY: Social Science Research Network.
- Chaney, Thomas (Sept. 2008). “Distorted Gravity: The Intensive and Extensive Margins of International Trade”. *American Economic Review* 98.4, pp. 1707–1721.
- Chemla, Gilles and Christopher A. Hennessy (Aug. 1, 2014). “Skin in the Game and Moral Hazard”. *The Journal of Finance* 69.4, pp. 1597–1641.
- Chemla, Gilles and Mario Milone (2018). “Oliver Hart - La Finance Vue à Travers La Théorie Des Contrats Incomplets”. *Les Grands Auteurs En Finance, 2ème Édition*. Éditions EMS.
- Coase, R.H. (1937). “The Nature of the Firm”. *Economica* 4.16, pp. 386–405.
- Costinot, Arnaud and Rodríguez-Clare Andrés (2014). “Trade Theory with Numbers: Quantifying the Consequences of Globalization”. *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Elsevier, pp. 197–261.

- Dewatripont, M. and E. Maskin (1995). “Credit and Efficiency in Centralized and Decentralized Economies”. *The Review of Economic Studies* 62.4, pp. 541–555.
- Diamond, Douglas (1984). “Financial Intermediation and Delegated Monitoring”. *The Review of Economic Studies* 51.3, pp. 393–414. JSTOR: 2297430.
- (1991). “Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt”. *Journal of Political Economy* 99.4, pp. 689–721.
- Diamond, Douglas W. and Philip H. Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity”. *Journal of Political Economy* 91.3, p. 401.
- Diamond, Douglas W. and Raghuram G. Rajan (2001). “Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking”. *Journal of Political Economy* 109, pp. 287–327.
- (2011). “Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes”. *Quarterly Journal of Economics* 126.2, pp. 557–591.
- Djankov, Simeon, Caroline Freund, and Cong S Pham (Jan. 8, 2010). “Trading on Time”. *The Review of Economics and Statistics* 92.1, pp. 166–173.
- Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer (May 1, 2007). “Private Credit in 129 Countries”. *Journal of Financial Economics* 84.2, pp. 299–329.
- Eaton, Jonathan and Samuel Kortum (2002). “Technology, Geography, and Trade”. *Econometrica* 70.5, pp. 1741–1779.
- Ellis, Andrew (Jan. 1, 2018). “Foundations for Optimal Inattention”. *Journal of Economic Theory* 173, pp. 56–94.
- Feenstra, Robert C. (2002). “Border Effects and the Gravity Equation: Consistent Methods for Estimation”. *Scottish Journal of Political Economy* 49.5, pp. 491–506.
- (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press.

- Feinberg, Yossi and Morton I. Kamien (2001). “Highway Robbery: Complementary Monopoly and the Hold-up Problem”. *International Journal of Industrial Organization* 19.10, pp. 1603–1621.
- Franck, Irene M. and David M. Brownstone (1986). *The Silk Road: A History*. Facts On File Inc. 356 pp.
- Frankel, Jeffrey A. and David Romer (1999). “Does Trade Cause Growth?” *American Economic Review* 89.3, pp. 379–399.
- Freixas, Xavier and Jean-Charles Rochet (2008). *Microeconomics of Banking*. The MIT Press.
- Froot, Kenneth A., David S Scharfstein, and Jeremy C. Stein (1993). “Risk Management: Coordinating Corporate Investment and Financing Policies”. *The Journal of Finance* 48, pp. 1629–1658.
- Froot, Kenneth A. and Jeremy C. Stein (1998). “Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach”. *Journal of Financial Economics* 47.1, pp. 55–82.
- Fuster, Andreas et al. (July 26, 2018). *Predictably Unequal? The Effects of Machine Learning on Credit Markets*. SSRN Scholarly Paper ID 3072038. Rochester, NY: Social Science Research Network.
- Gale, Douglas and Martin Hellwig (1985). “Incentive-Compatible Debt Contracts: The One-Period Problem”. *The Review of Economic Studies* 52.4, pp. 647–663. JSTOR: 2297737.
- Gardner, Roy et al. (2002). *Tolling the Rhine in 1254: Complementary Monopoly Revisited*. Working Paper.
- Gorton, Gary (2010). *Slapped by the Invisible Hand: The Panic of 2007*.
- Gorton, Gary and George Pennacchi (1990). “Financial Intermediaries and Liquidity Creation”. *The Journal of Finance* 45.1, pp. 49–71.
- (June 1, 1995). “Banks and Loan Sales Marketing Nonmarketable Assets”. *Journal of Monetary Economics* 35, pp. 389–411.
- Grossman, Sanford and Oliver Hart (1986). “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”. *Journal of Political Economy* 94.4, pp. 691–691.
- Hanson, Samuel et al. (2015). “Banks as Patient Fixed Income Investors”. *Journal of Financial Economics*.

- Hart, Oliver (1989). “An Economist’s Perspective on the Theory of the Firm”. *Columbia Law Review*.
- (2011). “Thinking about the Firm: A Review of Daniel Spulber’s The Theory of the Firm”. *Journal of Economic Literature* 49.1, pp. 101–113.
- Hart, Oliver D. and Dwight Jaffee (1974). “On the Application of Portfolio Theory to Depository Financial Intermediaries”. *The Review of Economic Studies* 41.1, pp. 129–147. JSTOR: 2296404.
- Hart, Oliver and John Moore (1990). “Property Rights and the Nature of the Firm”. *Journal of Political Economy* 98.6, pp. 1119–1158.
- Hartman-Glaser, Barney, Tomasz Piskorski, and Alexei Tchisty (2012). “Optimal Securitization with Moral Hazard”. *Journal of Financial Economics* 104.1, pp. 186–202.
- Head, Keith and Thierry Mayer (June 2010). “Illusory Border Effects”. *The Gravity Model in International Trade: Advances and Applications*. Ed. by Steven Brakamn and Peter van Bergeijk. Cambridge University Press.
- (Jan. 1, 2014). “Gravity Equations: Workhorse, Toolkit, and Cookbook”. *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Handbook of International Economics. Elsevier, pp. 131–195.
- Helpman, Elhanan, Marc Melitz, and Yona Rubinstein (2008). “Estimating Trade Flows: Trading Partners and Trading Volumes”. *The Quarterly Journal of Economics* 123.2, pp. 441–487.
- Hertzberg, Andrew, José María Liberti, and Daniel Paravisini (Apr. 1, 2011). “Public Information and Coordination: Evidence from a Credit Registry Expansion”. *The Journal of Finance* 66.2, pp. 379–412.
- Hirschman, Albert O. (1945). *National Power and the Structure of Foreign Trade*. University of California Press.
- Holmström, Bengt and Jean Tirole (2011). *Inside and Outside Liquidity*. MIT Press.
- Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba (2009). “The Trade Reducing Effects of Market Power in International Shipping”. *Journal of Development Economics* 89.1, pp. 84–97.

- Inderst, Roman and Holger M. Mueller (2006). “Informed Lending and Security Design”. *Journal of Finance* 61.5, pp. 2137–2162.
- (2007). “A Lender-Based Theory of Collateral”. *Journal of Financial Economics* 84.3, pp. 826–859.
- Innes, Robert D (Oct. 1, 1990). “Limited Liability and Incentive Contracting with Ex-Ante Action Choices”. *Journal of Economic Theory* 52.1, pp. 45–67.
- Jagtiani, Julapa and Catharine Lemieux (Apr. 5, 2018). *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the Lendingclub Consumer Platform*. SSRN Scholarly Paper ID 3178461. Rochester, NY: Social Science Research Network.
- Jappelli, Tullio and Marco Pagano (Oct. 1, 2002). “Information Sharing, Lending and Defaults: Cross-Country Evidence”. *Journal of Banking & Finance* 26.10, pp. 2017–2045.
- Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp (Mar. 1, 2016). “A Rational Theory of Mutual Funds’ Attention Allocation”. *Econometrica* 84.2, pp. 571–626.
- Karni, Edi and Subir K. Chakrabarti (1997). “Political Structure, Taxes and Trade”. *Journal of Public Economics* 64.2, pp. 241–258.
- Karolyi, Stephen Adam (Feb. 1, 2018). “Personal Lending Relationships”. *The Journal of Finance* 73.1, pp. 5–49.
- Kashyap, Anil K., Raghuram G. Rajan, and Jeremy C. Stein (2002). “Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking”. *The Journal of Finance* 57.1, pp. 33–73.
- Kim, Daesik and Anthony M Santomero (1988). “Risk in Banking and Capital Regulation”. *Journal of Finance* 43.5, pp. 1219–1233.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978). “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process”. *The Journal of Law & Economics* 21.2, pp. 297–326.
- Koehn, Michael and Anthony M Santomero (1980). “Regulation of Bank Capital and Portfolio Risk”. *Journal of Finance* 35, pp. 1235–1244.
- Krugman, Paul (1980). “Scale Economies, Product Differentiation, and the Pattern of Trade.” *American Economic Review* 70.5, pp. 950–959.

- Levitt, Steven D. and Christopher M. Snyder (1997). "Is No News Bad News? Information Transmission and the Role of "Early Warning" in the Principal-Agent Model". *The RAND Journal of Economics* 28.4, pp. 641–661. JSTOR: 2555780.
- Liberti, José María and Mitchell A. Petersen (Mar. 2017). *Information: Hard and Soft*. Working Paper.
- Limão, Nuno and Anthony J. Venables (2001). "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade". *The World Bank Economic Review* 15.3, pp. 451–479.
- Maggi, Giovanni (Jan. 1, 2014). "Chapter 6 - International Trade Agreements". *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Handbook of International Economics. Elsevier, pp. 317–390.
- Markowitz, Harry (1952). "Portfolio Selection". *The Journal of Finance* 7, pp. 77–91.
- Matejka, Filip and Alisdair McKay (Jan. 2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model". *American Economic Review* 105.1, pp. 272–298.
- Mayskaya, Tatiana (Aug. 4, 2017). *Dynamic Choice of Information Sources*. SSRN Scholarly Paper ID 2863605. Rochester, NY: Social Science Research Network.
- Melitz, Marc J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". *Econometrica* 71.6, pp. 1695–1725.
- Mester, Loretta J., Leonard I. Nakamura, and Micheline Renault (May 1, 2007). "Transactions Accounts and Loan Monitoring". *The Review of Financial Studies* 20.3, pp. 529–556.
- Meyer, Margaret A. (1991). "Learning from Coarse Information: Biased Contests and Career Profiles". *The Review of Economic Studies* 58.1, pp. 15–41. JSTOR: 2298043.
- Meyer, Margaret A. and Jeffrey Zwiebel (Mar. 2007). *Learning and Self-Reinforcing Behavior*. Working Paper.

- Miyagiwa, Kaz (Apr. 2009). *The Silk Road: Tax Competition among Nation States*. 18. Center for Research on Contemporary Economic Systems, Graduate School of Economics, Hitotsubashi University.
- Nimark, Kristoffer P. and Savitar Sundaresan (2018). *Inattention and Belief Polarization*. Working Paper.
- Norden, Lars and Martin Weber (Oct. 1, 2010). “Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers”. *The Review of Financial Studies* 23.10, pp. 3665–3699.
- Overman, Henry G., Stephen Redding, and Anthony J. Venables (2008). “The Economic Geography of Trade, Production, and Income: A Survey of Empirics”. *Handbook of International Trade*. Ed. by E. Kwan Choi and James Harrigan. Wiley-Blackwell, pp. 350–387.
- Padilla, A. Jorge and Marco Pagano (Dec. 1, 2000). “Sharing Default Information as a Borrower Discipline Device”. *European Economic Review* 44.10, pp. 1951–1980.
- Paravisini, Daniel and Antoinette Schoar (2015). *The Incentive Effect of Scores: Randomized Evidence from Credit Committees*. Working Paper 19303. National Bureau of Economic Research.
- Parlour, Christine A. and Guillaume Plantin (June 1, 2008). “Loan Sales and Relationship Banking”. *The Journal of Finance* 63.3, pp. 1291–1314.
- Pennacchi, George (2012). “Narrow Banking”. *Annual Review of Financial Economics* 4.0.
- Petersen, Mitchell A. and Raghuram Rajan (1994). “The Benefits of Lending Relationships: Evidence from Small Business Data”. *The Journal of Finance* 49.1, pp. 3–37.
- (May 1, 1995). “The Effect of Credit Market Competition on Lending Relationships”. *The Quarterly Journal of Economics* 110.2, pp. 407–443.
- (2002a). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *Journal of Finance* 57.6, pp. 2533–2570.
- (Dec. 1, 2002b). “Does Distance Still Matter? The Information Revolution in Small Business Lending”. *The Journal of Finance* 57.6, pp. 2533–2570.
- Philippon, Thomas (Aug. 2016). *The FinTech Opportunity*. Working Paper 22476. National Bureau of Economic Research.

- Puri, Manju, Jörg Rocholl, and Sascha Steffen (July 1, 2017). “What Do a Million Observations Have to Say about Loan Defaults? Opening the Black Box of Relationships”. *Journal of Financial Intermediation* 31, pp. 1–15.
- Pyle, David H. (1971). “On the Theory of Financial Intermediation”. *The Journal of Finance* 26.3, pp. 737–747.
- Raballand, Gael J. R. F. and Supee Teravaninthorn (Oct. 10, 2008). *Transport Prices and Costs in Africa : A Review of the International Corridors*. 46181. The World Bank, pp. 1–166.
- Raballand, Gaël, Antoine Kunth, and Richard Auty (Mar. 1, 2005). “Central Asia’s Transport Cost Burden and Its Impact on Trade”. *Economic Systems* 29.1, pp. 6–31.
- Raballand, Gael et al. (2012). “Why Cargo Dwell Time Matters in Trade”. *Economic Premise, World Bank* 81.
- Rajan, Raghuram (1992). “Insiders and Outsiders: The Choice between Informed and Arm’s-Length Debt”. *The Journal of Finance* 47.4, pp. 1367–1400.
- Rajan, Uday, Amit Seru, and Vikrant Vig (Feb. 1, 2015). “The Failure of Models That Predict Failure: Distance, Incentives, and Defaults”. *Journal of Financial Economics* 115.2, pp. 237–260.
- Redding, Stephen and Anthony J. Venables (Jan. 1, 2004). “Economic Geography and International Inequality”. *Journal of International Economics* 62.1, pp. 53–82.
- Repullo, Rafael (2005). “Liquidity , Risk Taking , and the Lender of Last Resort”. *International Journal of Central Banking* 1.2, pp. 47–80.
- Rochet, Jean-Charles (1992). “Capital Requirements and the Behaviour of Commercial Banks”. *European Economic Review* 36, pp. 1137–1170.
- Rose, Andrew K. (2004). “Do WTO Members Have More Liberal Trade Policy?” *Journal of International Economics* 63.2, pp. 209–235.
- Shleifer, Andrei and Robert W. Vishny (1992). “Liquidation Values and Debt Capacity: A Market Equilibrium Approach”. *Journal of Finance* 47.4, pp. 1343–1366.

- Shorrer, Ran I. (Dec. 2015). *Entropy and the Value of Informatio for Investors: The Prior-Free Implications*. Working Paper.
- Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Reprint, Liberty Classics, 1979.
- Stein, Jeremy C (2012). “Monetary Policy as Financial Stability Regulation”. *The Quarterly Journal of Economics* 127, pp. 57–95.
- Stein, Jeremy C. (2002). “Information Production and Capital Allocation: Decentralized versus Hierarchical Firms”. *The Journal of Finance* 57.5, pp. 1891–1921.
- Townsend, Robert M (Oct. 1, 1979). “Optimal Contracts and Competitive Markets with Costly State Verification”. *Journal of Economic Theory* 21.2, pp. 265–293.
- USAID (2004). *The Broad Economic Impact of Port Inefficiency: A Comparative Study of Two Ports*. Washington DC.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010). “Information Acquisition and Under-Diversification”. *The Review of Economic Studies* 77.2, pp. 779–805.
- Williamson, Oliver E. (1971). “The Vertical Integration of Production: Market Failure Considerations”. *The American Economic Review* 61.2, pp. 112–123.
- (1975). *Markets and Hierarchies: Analysis and Antitrust Implications : A Study in the Economics of Internal Organization*. Free Press. 312 pp.

Résumé

Cette thèse vise à déterminer et estimer les coûts de l'intermédiation et les externalités qu'elle génère.

L'avènement du big data et de l'IA remodèle l'intermédiation financière. Ses bénéfices en termes de vitesse et coûts sont visibles, mais son impact reste encore mal compris. Je montre que ces technologies peuvent accroître les frictions financières ainsi qu'augmenter les coûts de l'intermédiation financière.

La crise financière de 2008 a révélé l'importance des problèmes de liquidité. Je fournis un cadre d'analyse théorique pour comprendre les conséquences des régulations bancaires sur les problèmes de liquidité des banques. Il révèle un nouveau compromis entre les exigences en matière de fonds propres et de liquidités.

Enfin, les coûts et externalités de l'intermédiation sont étudiées dans le contexte du commerce international. Pour des raisons géographiques, certains pays sont dépendants du fait que leur biens transitent par des pays intermédiaires. Cette friction est analysée théoriquement et estimée empiriquement. Les pays qui semblent ne pas être affectés (tel que le Royaume Uni) souffrent des effets de distorsion des prix engendrés par ce problème.

Mots Clés

Théorie ; Régulation ; Banques ; Technologie ; Commerce International

Abstract

This thesis aims at determining and estimating costs of intermediation and the extend of the externalities it generates.

The rise of big data and AI is reshaping financial intermediation. Its benefits in terms of speed or cost efficiency may be easy to appreciate but its costs are far less understood. I show that technology can increase financial frictions as well as the cost of financial intermediation.

The financial crisis of 2008 has highlighted the importance of liquidity. I provide a theoretical framework to help understand the impact current regulatory tools may have on banks liquidity problems. The analysis sheds light on a new trade-off between capital and liquidity requirements.

Finally, costs and externalities of intermediation are studied in the context of international trade. Due to geographical constraints, some countries are intermediating goods in trade relationships. This trading friction is both theoretically analyzed and empirically estimated. Countries that are seemingly unaffected (e.g. The United Kingdoms) are in fact suffering from the distortionary effect it creates on global prices.

Keywords

Theory ; Regulation ; Banking ; Technology ; International Trade