NBER WORKING PAPER SERIES

INFORMATION: HARD AND SOFT

José María Liberti
Mitchell A. Petersen

Working Paper 25075
http://www.nber.org/papers/w25075

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

This is a significantly revised version of the working paper previously circulated as “Information: Hard and Soft” (2004) by Mitchell A. Petersen. The authors thank the editors Efraim Benmelech and Paolo Fulghieri for their patience and guidance in bringing this paper to fruition. We also thank Sumit Agarwal, Alan Berger, Richard Cantor, Bruce Carruthers, Beverly Clingan, Barry Cohen, Kent Daniel, Diane Del Guercio, Bob DeYoung, Joey Engelberg, Scott Frame, Andreas Fuster, Jon Garfinkel, Michael Faulkender, Andrew Karolyi, Juhani Linnainmaa, Tamim Majid, David Matsa, Gregor Matvos, Amit Seru, Philip Strahan, and conference participants at the University of Oregon, the Midwest Finance Conference, the University of South Carolina, and SFS Cavalcade for their suggestions and advice. The research assistance of Sang Kim, Austin Magee, and Mark Scovic is greatly appreciated. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information: Hard and Soft
José María Liberti and Mitchell A. Petersen
NBER Working Paper No. 25075
September 2018
JEL No. G2,G21,G3

ABSTRACT

Information is a fundamental component of all financial transactions and markets, but it can arrive in multiple forms. We define what is meant by hard and soft information and describe the relative advantages of each. Hard information is quantitative, easy to store and transmit in impersonal ways, and its information content is independent of its collection. As technology changes the way we collect, process, and communicate information, it changes the structure of markets, design of financial intermediaries, and the incentives to use or misuse information. We survey the literature to understand how these concepts influence the continued evolution of financial markets and institutions.

José María Liberti
DePaul University
1 E. Jackson Blvd.
DePaul Center 6131
Chicago
Illinois, 60604
jliberti@depaul.edu

Mitchell A. Petersen
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
mpetersen@northwestern.edu

I) Introduction.

Information is an essential component in all financial transactions and markets. A major purpose of financial markets and institutions in general is to collect, process, and transmit information. Given the importance of information and its transmission to the study of finance, as technology changes the way we communicate, it must fundamentally change financial markets, securities, and institutions, especially financial intermediaries. However, new technologies (i.e., those developed in the past fifty years) are more adept at transmitting and potentially processing information that is easily reduced to numbers. We call this hard information. Information that is difficult to completely summarize in a numeric score, where context matters, and where it is difficult to separate the collection of the information from its use is what we call soft information. Building upon the extensive literature on soft and hard information we examine the definition of these terms and their role in understanding financial markets and institutions.

The distinction between soft and hard information arose in the finance literature as a way to understand the evolving organization of lenders, although the theoretical ideas reach back much further. Banks have historically been a repository of information about borrowers' creditworthiness and the kinds of projects available to them. This information was collected over time through frequent and personal contacts between the borrower and the loan officer. Over time the banks built up a more complete picture of the borrower than was available from public records. This private information, whether hard or soft was valuable. The value, however, would be driven by the inability to replicate or transmit this information outside the bank (i.e., soft information).

The growth in the amount of numerical data available about borrowers and the ability to automate the credit decision transformed banking from an exclusively local and personal market to a national, competitive, and in some cases impersonal market. Some functions and decisions

which had resided inside the firm have been moved outside the firm due to a greater reliance on soft versus hard information. Information type (i.e. soft versus hard information) is an important characteristic of the lending environment and helps explain how the design of lending markets and institutions in which they operate have evolved.

The study of hard and soft information began in the banking literature. However, as technology progressed the role of soft or hard information in financial markets and institutions outside of banking and even outside of finance has grown. Not only did researchers use these concepts to examine a variety of financial markets and institutions (e.g., public equity markets, venture capital, municipal bonds, and real estate), but they examined how the type of information available to an institution helped determine the feasible and efficient organizational structures. Organizational constraints fed back into the kinds of information an institution could effectively use.

The purpose of the paper is to survey the literature on soft and hard information and to provide a review of what we know but also which questions remain unanswered. We describe what we believe to be the fundamental characteristics that define hard versus soft information in Section II. This provides a framework of how these terms have been used in the literature and which can be used to inform future work. We also discuss two historical examples of the hardening of soft information: the origin of credit ratings and the creation of the Center for Research in Security Prices. An institution's or market's decision to rely on hard or soft information is driven by what is available but also by the relative advantages of each. In Section III, we describe the main advantages of each type of information using the literature to provide examples and intuition. In Section IV we return to the roots of the soft and hard information literature. We start with a discussion of its foundation in the theoretical banking and organizational design literature, before

turning to efforts by the empirical literature to measure information type, directly and indirectly (e.g., by using geographic or organizational distance). This leads us to a discussion of the empirical challenge of designing incentives as a function of the type of information an institution uses. In the next section, we examine applications of soft and hard information outside of the banking literature. Specifically, we examine the lessons learned from the financial crisis as seen through the lens of information type. We also discuss the emerging work on FinTech which in many ways is the newest attempt by markets and firms to replace soft information with hard information. This section provides a guide to the future evolution of this literature as financial innovation and financial crisis are reoccurring themes in finance. Section VI concludes.

II) Defining Soft and Hard Information.

An initial challenge of using soft and hard information as useful constructs has been creating precise definitions. As the literature has expanded, the problem has not gotten easier. Thus, we will start with a brief description of the attributes of information that make it soft or hard. This description should be both consistent with much of the literature but also useful in framing research questions. Like many labels in finance (e.g., debt versus equity), there is no a clear dichotomy. Rather than two distinct classifications, we should think of a continuum along which information can be classified. Our interest is what characteristics of information, its collection, and use makes it hard or soft, and how these characteristics influence the structure of financial markets and institutions.

A) Characteristics of Soft and Hard Information.

1) Numbers versus text.

Hard information is almost always recorded as numbers. In finance we think of financial statements, the history of payments which were made on time, stock returns, and the quantity of

output as being hard information. Soft information is often communicated as text.[1] It includes opinions, ideas, rumors, economic projections, statements of management's future plans, and market commentary. The fact that hard information is quantitative means that it can easily be collected, stored, and transmitted electronically. This is why the advent of computers, large database programs, and networking has generate such growth in the use of production technologies that rely on hard information (e.g., quantitative lending, quantitative trading, and FinTech more generally).

2) The unimportance of context.

One can always create a numerical score from soft information, for example by creating an index of how honest a potential borrower is. This in and of itself doesn't make the information hard. Your interpretation of a 3 must be the same as mine. Thus, a second dimension of hard information is the unimportance of the context under which the information is collected. One can collect and code information and then transmit it to someone else. The meaning of the information depends only upon the information that is sent. It does not depend upon dimensions of the environment under which it was collected but which are not encoded in the data (Ijiri 1975). Thus, the receiver of the data knows all that the sender knows (or at least all that is relevant). With soft information, the context under which the information is collected and the collector of the information are part of the information. It is not possible to separate the two. This constrains the environments in which the data is collected and used. The environment has to be well-defined and predictable. In some cases, prior to entering the environments and collecting the data (e.g., talking

[1] Can't text be converted to numbers? Text files can obviously be translated into numbers; this is how they are stored and transmitted. Can't text files be processed electronically? Again, the answer has to be yes, conditional on what one means by processed. The ability of computers algorithms to process and generate speech (text) has improved dramatically since we first discussed soft and hard information. Whether it can be interpreted and coded into a numeric score (or scores) is a more difficult question. A numeric score can always be created, the issue is how much valuable information is lost in the process. We call this process the hardening of information and will discuss it below.

to a potential borrower), we know what variables we need to collect, what possible values those variable can take (i.e., a signal will be either good or bad; one, two, or three), and precisely how they will be used in the specific decision (values above 2 lead to a loan approval).[2] In other cases, prior to collecting the information, we are unsure what we might find or why it may be valuable until we after we collect the data.[3] Think of this knowledge as arising out of training and experience (Berger and Udell 2006). Later, when we are confronted with a decision, we recall the information collection process (e.g., the experience), and it is only then that is it apparent how the information we collected is useful. This is another characteristic of soft information. If we don't know what the information will be used for, or which parts of the information are relevant or useful, it is difficult to code and catalog it for future use.

The importance of context for soft information is related to the distinction in the contracting literature of whether a signal that is observable by outsiders is also verifiable by outsiders (Hart 1995; Aghion and Tirole 1997; Baker, Gibbons, and Murphy 2002). For a signal to be verifiable, the interpretation of the signal by the two contracting parties—and any third party who may be required to enforce the contract—must be the same. This is a characteristic of hard information. By contrast, soft information is private and not verifiable as it involves a personal assessment and depends upon the context, which is not easily captured and communicated. I can produce records

[2] A firm's sales revenue or their stock return is an example of such hard information. There is wide agreement as to what it means for a firm to have had sales of $10 million last year or the firm's stock price has risen by 10%. However, if we say the owner of the firm is trustworthy, there is less agreement about what this means and why it is important. Our definition of trustworthy may be different from yours and the context under which we evaluate their trustworthiness may be relevant.

[3] This distinction is reminiscent of the difference between the approach we take when we teach first-year graduate econometrics and the way empirical research is done in practice. In Econometrics 101, we assume we know the dependent variable, the independent variables, and the functional form. The only unknown is the precise value of the coefficients. In an actual research project, we have priors about the relationships between important economic concepts, but we don't know how to precisely measure the concept behind the dependent and independent variables nor the functional form. Only after collecting the data, and examining the preliminary results, that we understand how the variables are related. This leads us to modify our hypothesis and often requires the collection of additional data or alter our initial hypothesis and/or our interpretation of the data. The research process helps us see and understand the missing context.

that a borrower has paid their bills on time (hard information), but I cannot fully document that a borrower is honest as this relies on multi-dimensional observations and my personal assessment and standards. Following the organizational economics literature, hard and soft information can also be referred to as objective or subjective information.

3) Separation of information collection and decision-making.

The unimportance of context for hard information means it is possible to separate the collection of hard information and the decision-making based on that information. Adam Smith observed that the division of labor and specialization can create value in manufacturing; the same principal applies to the collection and processing of hard information. Knowing what information you are looking for and why it is valuable (i.e., what will it be used for) is essential if information collection is to be delegated. The collection of hard information does not even need to be personal. Hard information can be entered into a form without the assistance of or significant guidance from a human data collector (home mortgage mobile apps are an example). The data collector does not need to understand what decisions the data will be applied to. Soft information must be collected in person, and the information collector and the decision maker are often the same person.[4] This is the intuition in Stein (2002) for why smaller, less hierarchical firms are better able to use soft information in their decisions. This is also why relationship lending is built upon soft information (Berger and Udell 1995).

B) History of Soft and Hard Information

Historically, most information collection and decision-making was local and between individuals who were familiar with each other, and thus the distinction between hard and soft

[4] A typical example is the relationship-based loan officer. The loan officer has a history with the borrower and, based on a multitude of personal contacts, has built up an impression of the borrower's honesty, creditworthiness, and likelihood of defaulting. Based on this view of the borrower and the loan officer's experience, the loan is approved or denied. Uzzi and Lancaster (2003) provide detailed descriptions of the interaction between borrowers and loan officers.

information was not relevant. As technology made it feasible for financial transactions to occur between more distant and less familiar participants, the distinctions that we have discussing started to arise. We have implicitly been assuming that information type is static. Information is either hard or soft; it is not malleable. This simplification allows us to focus on the definition and advantages of each type of information. How discrete and immutable is information type? That is an empirical question. We can think of hard information as a numeric index, but soft information can and is converted into an index, but not without losing of information or context. Markets and individuals are constantly taking in soft (and hard) information and condensing (hardening) it into binary decisions: whether to fund a project, sell a stock, or make a loan. This does not create a meaningful loss of information if the decision is the final step and does not feed into later decisions.[5] Before moving on to the relative advantages of hard and soft information, we discuss two historic examples of the hardening of information and how the process changed the market or institutions that rely on this information. The examples are: the origin of credit ratings and the creation of the Center for Research in Security Prices (CRSP).

1) The Origin of Credit Ratings.

Credit ratings originated in the United States during the nineteenth century. Prior to this time, most trade among merchants was local. The extension of trade credit was common and merchants traditionally relied on soft information accumulated over time and through repeated personal interactions to make their credit decisions (Carruthers and Cohen 2001, Carruthers and Cohen 2009).[6] The development of communication and transportation technologies made it

[5] In Bikhchandani, Hirshleifer, and Welch's (1992) study of informational cascades, they model sequential decisions where agents see the (binary) decisions of prior agents but not the information upon which the information is made. This reduction (hardening) of information leads to agents ignoring their own (soft) information and following the crowd.

[6] The authors' description of trade credit markets during this period is strikingly similar to Nocera's (2013) description of the US consumer lending market of the 1950s.

possible to sell one's goods to a geographically much larger market. These were new customers with whom merchants had no prior personal experience, and thus their traditional approach to trade credit lending was not possible. These technological shocks created demand for new sources of information about creditworthiness that did not rely on direct personal connections, i.e., hard information. This led to the formation of firms which collected information about remote customers starting in the latter half of the 19th century.[7] These firms promised precise, standardized ratings that would allow merchants to avoid extending credit to distant customers who were not creditworthy.

The credit rating bureaus established local offices in major cities and relied on local merchants, lawyers, or bankers as the sources of their information. The input to the process was the same soft information that had previously been the basis of credit decisions.[8] The credit agencies used this information to create two credit scores which were sold to merchants: pecuniary strength (essentially net worth) and general credit (ability and willingness to repay). In this way, the agencies were able to take the soft information based on personal contacts and available to local merchants and provide it in a form that was useful to distant merchants. Merchants could make lending decisions based on this number, even though they had no contact with the potential customers or the data collectors. The standardization of the information in the form of the credit reports was a very early form of hard information and allowed the geographic reach of trade credit lenders to expand. It is also an example of how soft information can be hardened.

2) Creation of Center for Research in Security Prices (CRSP).

[7] The precursor to Dun and Bradstreet, the Mercantile Agency was founded in 1841 and the precursor to Standard and Poor's, the History of Railroads and Canals in the United States by Henry Poor was founded in 1860.
[8] "What went into the reports was a variable and unsystematic combination of facts and rumors about the firm, its owners, his personality, and family" (Cohen 1998).

The second historical example comes from the equity markets. The Center for Research in Securities Prices began as a database of monthly and then daily returns on all NYSE stocks in the early sixties: stereotypic hard information. There is rarely disagreement about what a return of four percent means. Prior to the construction of the CRSP databases, however, there was limited knowledge about what the returns on equities actually were, let alone what the determinants of equity returns were.[9] The existence of a comprehensive database containing the returns on all stocks unleashed a torrent of research into the determinants of both expected returns (e.g., factor models) and realized returns (e.g., event studies). It was now possible to carefully document what announcements or events influence stock prices (MacKinlay 1997). The dependent variable is a unidimensional index of value: the stock price (or changes in the stock price). The independent variables in this work are also coded into numeric values. Initially the coding was rudimentary: dividends increased, decreased, or did not change. Over time, the independent variable used to explain stock returns in the event study became more elaborate. However, they were always quantitative simplifications of the underlying events.

Although the event studies often found important determinants of stock prices, even when they focused on the individual days when seemingly large announcements were made, the fraction of cross-sectional variability that the models can explain is small (Roll 1984; Roll 1988). This

[9] CRSP began with a question from Louis Engel, a vice president at Merrill Lynch, Pierce, Fenner and Smith. He wanted to know what the long-run return on equities was. He turned to Professor Jim Lorie at the University of Chicago, who didn't know either but was willing to find out for them (for a $50,000 grant). The process of finding out led to the creation of the CRSP stock return database. The fact that neither the investment banks nor academic finance knew the answer to this question illustrates how far we have come in depending upon hard information such as stock returns. Professor Lorie described the state of research prior to CRSP in his 1965 Philadelphia talk "Until recently almost all of this work was by persons who knew a great deal about the stock market and very little about statistics. While this combination of knowledge and ignorance is not so likely to be sterile as the reverse-that is, statistical sophistication coupled with ignorance of the field of application-it nevertheless failed to produce much of value." In addition to CRSP, he talks about another new dataset: Compustat (sold by the Standard Statistics Corporation) which had 60 variables from firm's income statement and balance sheet.
http://www.crsp.com/research/james-lorie-recognized-importance-crsp-future-research

omission could be due to daily movements driven by the trading process (market micro-structure effects) or by the inadequacy of the explanatory variables. There are many forces that move stock price (rumors, news accounts, different interpretation of public releases) that are not easily and accurately converted to a numeric score. Although market participants capture this soft information and impound it into the hard information of stock prices, the academic models have had difficulty replicating the process.

III) Advantages and Disadvantages of Hard Information.

The choice between hard and soft information is driven by its availability and more importantly the relative costs of each. In this section, we describe the relative advantages of hard or soft information. The objective is both to explain why one kind of information is preferable in a given context, but also to more fully understand the definition of each.

A) Lower Costs of Production and Market Competition.

One of the major advantages of using hard information is the lower transactions costs (Frame, Srinivasan, and Woosley 2001). These come from several sources. First, by its nature, production technologies (such as loan origination) that depend upon hard information are easier to automate. The job of collection, and in some cases processing of information, can be delegated to lower-skilled workers or computers. Expensive labor can be replaced by cheaper labor or cheaper capital.

Hard information is more standardized. By construction it arrives in the same format and is processed in the same way for each application or transaction. The expertise of how to make the decision given the possible inputs is embedded into the decision rules or the computer code. This standardization introduces savings into the production process due to economies of scale. Once the computer system is designed and built to retrieve credit scores from the credit bureau and make

an approval decision, adding additional applications to the system has a small incremental cost. This is one reason why lending based on hard information (e.g., credit cards) has come to be dominated by large lenders much more so than traditional relationship lending (Cole, Goldberg, and White 2004; Berger, Miller, Petersen, Rajan, and Stein 2005; Berger and Black 2011). These potential cost savings have created economic pressure to find ways to automate small loans to firms or individual as a large fixed cost can make these loans prohibitively expensive.[10]

The greater reliance on hard information may also increase the competitiveness of financial markets.[11] First, the standardization of information and the resulting lower transactions costs can expand the size of the market by increasing the number of suppliers who can profitably offer such loans or services. In addition to expanding the number of suppliers in a given market, a reliance on hard information can also increase the geographic reach and competitive impact on existing suppliers. The evolution of the mortgage and signature loan (now called the credit card) market is an example.[12] In the 1950s, the market was local and based on soft information obtained through personal contact. It is now national and based on hard information often obtained through impersonal contact. This has led to a wider availability of and arguably cheaper capital for the middle class (Nocera 2013).

The nature of the information may also increase the competitiveness of the markets. Once information is systematized and easy to communicate (hard), it also becomes more difficult to contain. In the early years of the credit reporting agencies (e.g., Bradstreet or R. G. Dun), only a summary of the information the agencies had on borrowers was published in their quarterly books.

[10] For small business loans, the size of the fees is independent of the size of the loan. Thus, the percentage fee declines with loan size (Petersen and Rajan 1994).
[11] The causation can also run in the opposite direction. Greater competition, which can arising from deregulation for example, increases the pressure to lower costs and thus transform the production process to depend more on hard information.
[12] Subprime mortgage loans are less standardized and more informationally sensitive than normal mortgages because sometimes borrowers are not able to provide full disclosure of their income (Mayer, Pence, and Sherlund 2009).

This disclosed information was quantitative and easy to compare and communicate. For an additional fee, subscribers could visit the office of the agencies to view a detailed report on a potential customer. The credit rating bureaus were either unable or unwilling to quantify and include all of the soft information they held into their reported credit scores. Interestingly, information in these private reports was better at predicting bad outcomes (business failures) than the published credit ratings (Carruthers and Cohen 2001). By keeping the information difficult to replicate and transmit, by maintaining its softness, the credit reporting agencies hoped to maintain their control over the information and thus extract greater rents from the information they had collected. Once information is hard, providers have difficulty preventing one customer from passing it to additional customers who can then capture the information's full value. Information that is hard can be understood independent of the collector and the context under which it was collected. If the collector is not necessary once the user has the data, this makes charging high rates for the information more difficult.

B) Durability of Information.

The durability of information is also greater when it is hard. The fact that it is easily stored means that the cost of maintaining it for future decisions is low. The fact that the information can be interpreted without context means that it is possible to pass it along to individuals in different parts of an organization (Stein 2002). Individuals or even firms no longer need to be part of the data collection process to be part of the decision-making process. This is especially important if the people involved in data collection are not expected to be around in the future. This frees the decision process from constraints of space (distance) and time. Given the increased turnover in many finance professions (loan officers or investment bankers), the movement toward hard

information seems inevitable.[13] As described in Crane and Eccles (1988), junior investment bankers used to rise through the bank as junior employees of their clients simultaneously rose through the ranks at their own firms. By the time junior bankers became senior bankers, they had developed a relationship with the people who were now in senior positions in the bank's clients. There is no need to rely on formal records (hard information) in the presence of these long-term relationships. However, if bankers turn over more frequently, new bankers must rely on the records left behind by the previous bankers (Morrison and Wilhelm 2007). This creates a greater reliance on hard information.

C) Lost Information.

Part of the reason that hard information is less costly to communicate is that there is less of it. The replacement of soft with hard information inevitably results in a loss of information (as when an analog signal is converted to a digital signal). This is why it is possible to use a smaller bandwidth to transmit the information. As an example, compare two methods of making a loan approval decision. First is the stereotypic credit scoring decision, in which a finite number of quantitative variables are weighted and summed to obtain a credit score. The loan is approved if the value of the score is above a critical value. Now compare this to the traditional relationship approach to lending. After spending several hours discussing the borrower's investment plans and using the loan officer's years of experience with the borrower and knowledge of the local business environment, a decision is rendered. Both decision-making methods lead to a binary approval or rejection decision, but the first requires less information as inputs to the decision.

[13] Karolyi (2017) finds that the relationship lies with the individuals not the firms. After exogenous changes in leadership (the death of retirement of a CEO), firms are significantly more likely to switch to lenders with whom the new CEO has a relationship (see also Degryse, Liberti, Mosk, and Ongena 2013). This is one reason why firms that rely on soft information in securing debt capital care about the fragility of the banks from which they borrow (Schwert 2017).

The reduction of information is never good, as long as processing costs are zero. However, decision makers (e.g., the loan approval committee of a bank) has limited time and attention to devote to each decision.[14] To prevent information overload, decision makers need the information to be boiled down to what is important.[15] The larger the organization and the higher one goes in the organization, the more the information needs to be concentrated or decision-making authority delegated. The question is how important is the lost information. The concern about small firms' and individuals' access to capital in the presence of bank consolidation and the growing use of credit scoring type lending decisions is driven by this question (O'Neil 2016). If there are borrowers that are really good credit risks, but they look bad on paper (i.e., when only hard information is considered), then such borrowers would be incorrectly denied credit. The empirical evidence thus far is mixed. It is clear that some small borrowers are dislocated by their banks when the banks merge, but there is also evidence that existing and new small banks may fill the gap (Berger, Miller, Petersen, Rajan, and Stein 2005; Berger, Goulding, and Rice 2014; DeYoung, Gron, Torna, and Winton 2015; Berger, Bouwman, and Kim 2017).

D) Gaming the System.

Accounting numbers, such as firm's income statement and balance sheet, are a classic example of hard information. The information is all quantitative, easy to store and transmit electronically, and there is relatively uniform agreement about what numbers like revenues and costs mean. This is why quantitative decisions from asset allocation to credit approval rely on these

[14] Using a randomized control trial, Paravisini and Schoar (2015) evaluate the adoption of credit scores in a small business lending setting. They find that using credit scores improves the productivity of credit committees (e.g., less time is spent on each file).

[15] Friedman (1990) argues this is one advantage of a market versus a planned economy. He argues that all of the information that is relevant to a consumer or producer about the relative supply of a good is contained in the price. Thus, it is not necessary for a supplier to know whether the price has risen because demand has risen or supply has fallen. The supplier only needs to know that the price has risen, and this will dictate his decision of how much to increase production. His description of the market is a classic hard information environment.

numbers. Newspaper accounts of accounting manipulation and the size of the credit rating manuals make it clear that these decisions are not simply a function of the numbers the firms disclose. This raises another cost of using hard information (e.g., automated or delegated decisions methods): who controls the information which is fed into the decision-making process.

The discussion thus far has focused on the decision maker (e.g., the bank making a loan decision), not the target of the decisions (e.g., a loan applicant). By choosing to use hard versus soft information, the decision maker is trading off the lower cost of collecting and processing the information with potential loss in accuracy of the information upon which they are basing their decisions. The way a decision is made, and the type of information upon which the decision depends, will also influence the actions of the target of the decision.

The behavioral response of borrowers (or other targets of the decision) places restrictions on how decisions based on hard information can be made. Having a decision depend only upon the numbers and a transparent decision rule can work only if the cost of manipulating the numbers is sufficiently high relative to the benefit of the preferred outcome.[16] If a firm can raise its reported assets or sales by a small amount for a small cost, and this will raise its credit rating and lower its

[16] In the financial crisis of 2008, a large number of investment grade securities defaulted. The magnitude of the defaults suggested there was a problem with the rating process (see Benmelech and Dlugosz 2009a; Benmelech and Dlugosz 2009b). Observers in industry, academics, and government suggested possible sources of the problem and potential solutions. What is intriguing is the defaults experience was very different in the corporate bond market (debt of operating companies) compared to the structured finance market (e.g., RMBS). Defaults in the corporate bond market spiked in 2009, but the peak is not drastically different than the peak in prior recessions (see Vazza and Kraemer 2016, Chart 1). The peak in defaults in the structured finance in 2009 is dramatically larger (see South and Gurwitz 2015, Chart 1). Although the collapse of the housing market hit the structured finance securities harder, this suggests that a part of the problem with the rating process resides uniquely in the structured finance segment of the market. For an operating company, a low cost of capital is an advantage but not its only or predominant source of competitive advantage. For a securitization structure, a lower cost of capital is one of its few source of "competitive advantage." A bank might change which mortgages are placed into a securitization if this increased the faction of the securitization rated AAA and thus lowered the cost of capital. An auto-manufacturing firm is unlikely to close plants or close down a division solely to get a higher credit rating. The costs of altering the business to improve a credit score are higher and the benefits are (relatively) lower for an operating firm. This may be why we saw relatively fewer defaults in the corporate bond sector relative to the securitized sector. This issue prompted the credit rating agencies to consider different rating scales for structured finance versus corporate debt (Kimball and Cantor 2008).

cost of capital sufficiently, it has an incentive to inflate its reported assets or sales.[17] The rules cannot be a direct and transparent functions of the hard numbers if the hard numbers are under the discretionary control of the market participants. In this case, the decision maker has an incentive to make the decision a fuzzy and opaque function of the inputs. The line between an AA and an A rating can be kept secret or additional sources of soft information can be included.[18] In practice, ratings models that try to explain the ratings as a function of the firm's financial numbers have R^2 appreciably below 100 percent.

E) The Role of Discretion.

Hard information reduces the information that is used, but equally importantly delegates the decision-making authority. The individual collecting the data does not make the decision. This has been delegated to a higher up or to an algorithm (whose author is divorced from the target of the decision). This should eliminate discretion, and this can be a positive or a negative. Relationships are useful as a way to elicit information that is not available in the numbers. Relationships have additional, non-informational, dimensions as well. Relationships generate a sense of mutual obligation (reciprocity). You help me out and I want to help you out (Uzzi 1999). Thus, when a loan officer is evaluating a potential loan from a long-term borrower, they can use their discretion to more accurately evaluate the borrower's current credit quality as well as changes in the likelihood of repayment. A borrower is not defaulting on an obligation to an unknown

[17] Hu, Huang, and Simonov (2017) see the same behavior in the market for individual loans. The theoretical importance of nonlinearities in the mapping of inputs (hard information) to outputs (decisions) is discussed in Jensen (2003). In his examples, the incentives to misstate one's information are smaller if the payoff function is linear. Small changes in the reported information have only small changes in the manager's payoff.

[18] There may also be strategic reasons to avoid a transparent mapping between the numbers and the credit rating. The business model of credit rating agencies relies on market participants being unable to replicate the ratings at lower cost than the agency. If the mapping were a direct function of easily accessible inputs (e.g., the income statement and balance sheet) and nothing else, some clever assistant finance or accounting professor would figure out the function. This is one reason that the early credit reporting agencies released only a fraction of their information publicly in the form of a credit score. For additional fees, users could review a more complete report (Carruthers and Cohen 2001).

faceless financial institution, in these cases, but to someone with whom they have worked for years.[19] This is the positive side of the relationship that can be lost with decisions based on hard information.

Loan officers can also use their discretion to put a thumb on the scale and influence a loan decision for their own benefit. A number of academic papers have documented that loan officers do use their discretion, and in the documented cases, the discretion does not improve the quality the decision.[20] The challenge is one of incentives. The loan officers are not lending their own capital, but the bank's. The bank manager or shareholder must trade-off the value of the loan officer using their soft information (better quality decision and lower transactions costs) against the misaligned incentives between the loan officer and the bank. The advantage of hard information is it can remove the loan officer's discretion. The relevant variables and the mapping from the variables to the decision is beyond the control of the loan officer in these cases.[21]

IV) Traditional Banking and the Organizational Design of Lending.

The evolution of financial markets over the past forty years has been in part a replacement of soft information with hard information as the basis for financial transactions. The full

[19] Guiso, Sapienza and Zingales (2013) find that borrowers feel less obligated to repay an underwater mortgage if the mortgage has been sold in the marketplace.

[20] Brown, Schaller, Westerfeld, and Heusler (2012) find that loan officers use discretion to smooth credit, but there is limited information in discretionary changes. Degryse, Liberti, Mosk, and Ongena (2013) provide evidence that soft information helps predict defaults over public information (e.g. financial statements), but discretionary actions do not predict default. Gropp, Gruendl, and Guettler (2012) show that the use of discretion by loan officers does not affect the performance of the bank portfolio. Puri, Rocholl, and Steffen (2011) document the widespread use of discretion inside a German savings bank but find no evidence that loans approved base on discretion perform differently than those that do not use discretion. Cerqueiro, Degryse, and Ongena (2011) find that discretion seems to be important in the pricing of loans, but plays only a minor role in the decision to lend.

[21] This turns out to be an imperfect solution when the loan officers has an incentive to manipulate the inputs, just as the borrower might. The loan officers in Berg, Puri, and Rocholl (2016) work for a bank that uses an internal credit score to evaluate loans. They show that loan officers repeatedly enter new values of the variables into the system until a loan is approved. Not only are they able to get loans approved that were originally rejected, but they also learn what the model's cut offs are and thus what is required to get a loan approved. These results suggest that even hard information decision-making algorithms which are transparent and depend upon data subject to the control of either participant (local decision maker or the target of the decision) are subject to the Lucas critique (see the Gaming the System discussion above).

ramifications of this transformation are not yet fully apparent, and as we discussed above, there are both advantages and disadvantages of this transformation. In this section, we describe the evolution of soft information since its theoretical origins, the application of the concept of information type in the traditional banking literature, and its implications for the organizational design of lending by financial intermediaries.

A) Beginnings: Theoretical Literature.

The finance literature has been exploring the distinction between soft and hard information for several decades now and our understanding has evolved since the early years. The distinction was not always explicitly stated, and even when it has been, the definition was not complete, formally treated, or consistent across applications.

One origin of soft and hard information traces back to the theoretical financial intermediation literature and distinction it drew between the role of banks (or other private lenders) versus the public bond markets. A key distinction was the superior ability of banks to collect and process information (Diamond 1984; Diamond 1991; Ramakrishnan and Thakor 1984; Allen, Carletti, and Gu 2015). This explained why many opaque firms relied exclusively on banks. The public debt markets, however, with the help of rating agencies, have the same job description: to evaluate the credit quality of firms (Ederington and Goh 1998). The difference is the type of information each specializes in collecting and processing.

The public bond markets and the rating agencies collect financial disclosures, accounting reports, and default histories. These are sources of hard information. They can all be reduced to a series of numbers. Banks, on the other hand, especially as described by the lending relationship literature, collect information that is neither initially available in hard numbers (the ability of the managers, their honesty, the way they react under pressure), nor easily or accurately reducible to

a numerical score. Even if reduced to a numerical score, the interpretation of the information may be judgmental and include a discretionary component (Cole, Goldberg, and White 2004; Hertzberg, Liberti, and Paravisini 2010). Once the relationship is established, even then this information is not hard. The firm is still unable to communicate this information to the broader lending markets and thus negotiate a lower loan rate from its bank (Petersen and Rajan 1994).

Originally, finance scholars borrowed the concept of soft information from organizational economics and the theoretical literature on decisions making in organizations. One feature of those initial models was that interests of parties were imperfectly aligned. This misalignment created incentives for individuals to distort the collection and transmission of information that influenced decisions to their advantage (Milgrom and Roberts 1988).[22] The fact that information needs to be transmitted to a superior that ultimately has final authority in the decision-making process, in our case a loan officer that transmits soft information to their supervisors, led the traditional banking literature to analyze the role of organizational form in lending.

B) The Role of Organizational Form in Lending.

In many industries, both large and small firms coexist. One might think that a dominant production technology would lead to a uniform firm size. However, if the information collection, processing, and communication is fundamentally important to the production process (e.g., banking, drug research, or film production, Goetzmann, Ravid, and Sverdlove 2013), then firms may specialize in different sectors of the market depending upon the type of information (hard or soft) that is used in their production process. Some firms may specialize in production processes

[22] There are a variety of possible costs embedded in the transmission of information in an organization. Theories of costly communication, where soft information may be more costly to communicate across hierarchies (Becker and Murphy 1992; Radner 1993; and Bolton and Dewatripont 1994); theories of loss of incentives to collect, process and use soft information as in Aghion and Tirole (1997) due to the anticipation of being overruled by one's superior, and strategic manipulation of information as in Crawford and Sobel (1982) and Dessein (2002). In these theories, those who send the information will make it noisier and less verifiable if their preferences are not aligned with those who are receiving it and, ultimately, have the final authority to make the decision.

based on soft information, and others in a production process based on hard information. Stein (2002) argues that larger, more hierarchical firms, where the decision maker is further from the information collector, are more likely to use production technologies that rely on hard information (Brickley, Linck, and Smith 2003). These organizational diseconomies suggest that large banks are expected to be less efficient at making relationship loans—that is, loans that depend upon soft information. Information in a large bank is potentially collected by one individual or group, and a decision is made by another. These decisions require information that can efficiently be transmitted across physical or organizational distances. The information must also have a uniform interpretation that does not depend upon the context under which it was collected. Large banks are more likely to have multiple layers of management. They are hierarchical or centralized opposed to flat or decentralized organizations. Thus, the oversight of loans in this context implies that larger banks rely relatively more on hard information (Berger, Miller, Petersen, Rajan and Stein 2005; Qian, Strahan, and Yang 2015; Liberti 2017).

The literature on organizational form in financial institutions has exploited the distinction between hard and soft information to help explain the scope of the firm, but these ideas appear in the literature much earlier. As noted in Section II-A-3, a key feature that made scholars interested in exploring issues of soft information in the traditional banking is that the process of collection and decision-making are separated, thus transmission is key. Since Coase (1937), the idea of allocating control and decision-making within organizations has been a core principal of the theory of the firm. The allocation of control shapes the incentives of agents working in the organizations. In seminal papers, Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995) define allocation of control from the ownership of a tangible asset. In the case of financial institutions (or other firms whose production process depends upon intellectual property or information that is not

easy to transmit), the critical resource or asset is intangible in nature: the access to information, especially soft information that needs to be communicated to the decision maker.[23]

As financial institutions have become larger, more globalized, and more complex they face a tradeoff between benefits arising from economies of scale and costs of inferior organizational designs. This has led to a debate in the banking literature over whether a decentralized organizational structure is better or worse than a centralized one in terms of providing the right incentives to loan officers to collect, process, transmit, and use soft information. The discussion has centered on how the informational distance between the decision maker and the loan officers shapes the nature of information acquisition and, therefore, the types of activities performed inside or outside the financial institution. As innovations in communication technology have reduced the cost of accessing information at a distance, this has also changed the competitive landscape of banking (see also the discussion of Fintech in Section V-C). For example, Liberti, Seru and Vig (2017) examine the introduction of a credit registry in Argentina. They find that it led to an improvement in the allocation of credit to borrowers for whom there was now more public hard information available, but it also changed the internal organization of the bank. We next turn to the empirical literature that has explored these issues.

C) From Geographical to Hierarchical Distance.

Based on the theoretical predictions on the challenge of transmitting some types of information (i.e., soft information), the empirical literature has attempted to document the importance of distance. An important branch of the traditional banking literature shows that geographical distance affects lending decisions (Petersen and Rajan 2002; Degryse and Ongena

[23] Rajan and Zingales (1998) argue that ownership is not the only way to allocate power in an organization. Another and in some cases a better way is through access. Access is the ability to work with or use a critical resource, not necessarily a physical resource that can be owned. In financial institutions (and increasingly in other firms), this asset is often information.

2005; Mian 2006; DeYoung, Glennon, and Nigro 2008; Agarwal and Hauswald 2010).[24] The literature has interpreted this finding largely in terms of the difficulty of transmitting soft information. Despite its prominence, this interpretation is largely based on the observed correlation between loan characteristics and distance.

A classic example of a first generation paper using geographic distance is Berger, Miller, Petersen, Rajan, and Stein (2005). Consistent with Stein (2002), they find that larger banks are more likely to lend to more distant customers (a greater physical distance between a firm and its bank) and communicate with borrowers more impersonally (by mail or phone opposed to face to face). They also find that relationships between a firm and its banks are less durable and less exclusive when the banks are larger. Most importantly, they find that firms that are forced to choose a larger bank than they would prefer a small bank (i.e., informationally opaque firms) are credit rationed. When informationally opaque firms have the choice of which size bank to borrow from, they choose to borrow from smaller banks. The correct matching alleviates much of the credit rationing.

Second generation papers focused on the organizational structure of financial intermediaries. The lower costs of producing hard information can depend on more than just the nature of the information. It may also depend upon the organizational design of financial institutions. Lenders who are larger in size and hierarchically organized benefit from economies

[24] Although these papers are on geographical distance, they are different in nature. Petersen and Rajan (2002) document that distance between lenders and borrowers increased due to improvements in information technology. Degryse and Ongena (2005) study the relationship the competitiveness of the lending market and the distance between the borrower, their lender, and other potentially competitors (banks). Mian (2006) suggests that greater distance not only decreases the incentives of a loan officer to collect soft information, but also makes it more costly to produce and communicate soft information. DeYoung, Glennon, and Nigro (2008) document the relationship between the use of hard information using credit scoring technologies and increases in borrower-lender distances. Finally, Agarwal and Hauswald (2010) study the effects of distance on the acquisition and use of private information in informationally opaque credit markets. They show that borrower proximity facilitates the collection of soft information, which is reflected in the bank's internal credit assessment.

of scale in using hard information but can find it more costly to transmit soft information. This will cause them to place a greater reliance on hard information. Nevertheless, large banks may try and mimic the organizational structure of small banks in order be to more efficient in collecting soft information and, therefore, improve their ability to compete against smaller banks. For example, Liberti and Mian (2009) use hierarchical distance between loan officers and their superiors to study the causal impact of hierarchical structures on the relative importance of hard versus soft information in the credit approval decisions inside a large financial institution. The authors find that greater hierarchical distance is associated with less reliance on soft information and more on hard information. They also find that personal interaction between loan officers and the superiors approving the loans helps mitigate the effects of hierarchical distance on information use and minimizes the loss of soft information (see also Qian, Strahan, and Yang 2015).

One natural question is why delegation of decision rights is a solution to the problem of transmission of soft information. There are several potential explanations for this to be the case. An incomplete list includes increase capacity in a limited resources environment, expertise, communication costs, and ex-ante incentives, among others. Assigning a task to another person adds capacity to carry out the task relying on the comparative advantage of that specific individual. A loan officer collecting and processing soft information is a time-consuming activity, and given the limited resources of her superior, she may decide to delegate certain tasks. There may be efficiency gains from using a loan officer with the ability and skill set to collect soft information (Geanakoplos and Milgrom 1991).

Delegation of decisions may also arise because of the difficulties of communicating specific information to the superior making communication costs high as in Jensen and Meckling (1992). A key tradeoff in the organizational design of lending occurs between efficient

communication and the cost of collecting soft information. Garicano (2000) explores the acquisition of knowledge by the creation of knowledge-based hierarchies in a more general setting. In these hierarchies, certain individuals solve the easiest problems and more difficult problems are solved by specialized supervisors. In our banking setting, loan officers may have the easier task of collecting soft information while supervisors use the information. There is another strand of literature that focuses on the ex-ante incentives (Aghion and Tirole 1997, Crawford and Sobel 1982, Dessein 2002). For example, Aghion and Tirole (1997) and Stein (2002) argue that large hierarchical organizational structures inhibit the ex-ante incentives to collect and use soft information. This decrease in incentives occurs because those in charge of collecting soft information cannot act on it, and instead have to send the information to their superiors. The nature of soft information means it may be overruled or ignored, thus dampening the incentives to collect soft information. Liberti (2017) provides support for the loan officers' incentives view by showing that loan officers who receive relatively more authority put more effort into collecting and using soft information.

D) Implications of Relying on Soft Information

Consolidation of financial institutions may have a negative impact on small business lending due to the potential loss of soft information and the incentives to collect it going forward.[25] Small banks are superior at relationship lending using soft information because of their fewer layers of management making it easier to communicate and use such information. Large banks can

[25] Starting in the early eighties the number of banks in the US has declined by over fifty percent, with most of the fall occurring in the first decade (Petersen and Rajan 2002, Figure 4; Berger and Bouwman 2016, Figure 8.1). The decline in the total number of banks is completely driven by the decline of small banks defined by those with gross total assets less than $1B. The number of large banks has grown. The decline in small banks is driven in part by the technology and the shift to hard information, but also by deregulation (Strahan and Kroszner 1999). The growing reliance on hard information and automated decision-making, and the associated cost savings, created pressure to reduce regulations on bank expansion. In turn, as regulatory restrictions diminished, this raised the value of capturing cost savings by shifting to production processes which rely on hard information.

simulate the managerial environments of small banks, in order to minimize the negative impact of losing information (Berger, Miller, Petersen, Rajan, and Stein 2005). The results in Liberti (2017) also highlight how a large bank may be able to replicate the organizational structure of a small bank by delegating decision-making authority to the lower layers of the organizations.

Along these lines, Agarwal and Hauswald (2016) provide direct evidence that the findings on distance and loan characteristics in the existing literature are really due to the difficulty of transmitting soft information. In other words, they provide evidence that a bank endogenously responds to information transmission problems by effectively delegating more authority to loan officers. Skrastins and Vig (2018) also find evidence that increasing the hierarchical structure of a branch decreases the ability of the loan officers to produce soft information, leading to an increased standardization of the information collected for each loan.

The empirical literature has found that firms' access to capital depends upon how informationally transparent the firms are or how much hard information the financial markets have about the firms. We expect small firms to face greater credit rationing because of the limited hard information available about them. This is why they are more reliant on banks that are better at extracting and using soft information. However, when we look only at small firms, we still find that a firm's access to credit is a function of how much information is available to the financial markets not just the bank. Firms that are more informationally transparent, for example if they maintain formalized records, find that they have a higher probability of their loans being approved (Petersen and Rajan 2002).

For publicly traded firms, the amount of hard information available about the firm is much greater. However, even for publicly traded firms, the existence of information that is easy to access and evaluate on their likelihood of default—such as a credit rating—appears to increase their

access to debt capital (Petersen and Faulkender 2006). Controlling for traditional determinants of capital structure (e.g., taxes, asset tangibility, and growth opportunities), firms with a debt rating have 35 percent more debt than otherwise identical firms.

V) Applications of Soft and Hard Information beyond Traditional Banking.

Although much of the initial research on the implications of information type focused on the banking and lending market, the literature has expanded beyond that. In this section we discuss three additional areas where researchers have applied these concepts: distance research outside of banking, the financial crisis, and financial technology (Fintech).

A) Distance Research Outside of Banking.

Based on the research in the banking, distance is related to information type. Hard information can be transmitted across distance without loss of content; whereas soft information cannot. This raises the question as to which financial markets must be geographically close and which do not need to be. This analysis helps us understand what kind of information undergirds each market. The finance literature has studied distance in a variety of other economic settings and markets including: the public equity markets (Grinblatt and Keloharju 2001; Hong, Kubick, and Stein 2005; Malloy 2005), the municipal bond market (Butler 2008), the venture capital market (Lerner 1995), the real estate market (Garmaise and Moskowitz 2004; Granja, Matvos, and Seru 2017), the allocation capital between divisions within a firm (Landier, Nair, and Wulf 2009), and the impact of the organization design of conglomerates on their productivity (Seru 2014).[26] In part due to credit ratings, the corporate bond market is national or international. Even though municipal bonds (tax-exempt bonds issued by state and local government entities) are also rated, Butler (2008)

[26] Even in markets that we think are dominated by hard information and thus, where we would expect distance not to be relevant, research has sometimes found a preference for local investments. Mutual fund managers tend to hold a higher concentration in shares of local firms since access to soft information of local firms is cheaper (Coval and Moskowitz 1999; Coval and Moskowitz 2001). The effect is strongest in small and highly levered firms.

finds that the underwriting market is highly local (80% of municipal bonds are underwritten by investment banks with a local office). Unlike small business loans, the underwriters of municipal bonds do not hold the securities and so do not have incentives to monitor borrowers post issuance. The local underwriters have been able to credibly communicate to investors that their soft information is valuable and certify the bond's quality. Local underwriters are able to sell municipal bonds for higher prices and these results are strongest when the ratings signal is weakest (i.e., bonds with low ratings and unrated bonds).[27]

One of the most opaque financial sectors is the equity of new and private firms. There is little to no hard information available about such investments, in part because they have no track record and in part because they are often in emergent industries. Due to their "…detailed knowledge of the firm they finance, (venture capitalists) can provide financing to young businesses that otherwise would not receive external financing" (Lerner 1995). The venture capitalist acquires soft information by serving on the firm's boards, making frequent visits to the firm, and meeting with customers and suppliers. Since this is costly in terms of time, distance matters. A venture capitalist is twice as likely to sit on the board if their office is within five miles of the firm compared to 500 miles (Lerner 1995).

The discussions thus far has focused on the role of information type in explaining external distances. Distances inside a firm may be relevant as well. Landier, Nair, and Wulf (2009) explore how the distance between divisions and headquarters within the same conglomerate may have an impact on corporate decision-making. Managers are more likely to lay off employees or divest divisions that are more distant from headquarters. Although the authors argue this could be driven

[27] If the local underwriters have soft information that non-local underwriters do not have and they can thus sell the bonds at higher prices, they should be able to extract larger fees. Oddly, they do not. Local underwriters charge lower fees relative to non-local underwriting suggesting local competition is limiting their pricing power.

by a greater affinity for the people management interacts with most often, they also find the effect is stronger in environments that rely on soft information.[28] The problem with interpreting this empirical result is the choice of locations is likely to be endogenous, making it difficult to establish causality. Instead, Giroud (2013) has a clever way of solving the endogeneity problem. He studies the impact on plant-level investment and productivity when the headquarters is either close or far from the plants, arguing that travel time is a better proxy for monitoring than geographical distance, especially when the inputs to the management's decision are soft information.[29] He exploits the introduction of new airline routes as a source of exogenous variation to proximity and measures the causal impact of distance on plant-level investment and productivity. This makes sense if the information the managers need to acquire is soft and thus obtained only by visiting the plants more frequently.

B) Financial Crisis: The Role of Information Type and Incentives.

A reliance on hard information was an essential factor in enabling the growth of securitization and the ensuing expansion of mortgage lending. If lenders have information that they could not pass along (e.g., soft information) and they use this information to determine which loans to sell, the securitization market can break down (Stiglitz and Weiss 1981; DeMarzo 2005). The reliance on hard information in the mortgage market reduced transactions costs, expanded access to capital, and diversify regional risk (Ranieri 1997; Demsetz and Strahan 1997; Allen and Carletti 2006; Loutskina and Strahan 2011). The reliance on hard information also had a dark side

[28] They use the measure of distance between banks and borrowers from Petersen and Rajan (2002) to classify whether industries are hard- or soft-information intensive. Industries where distance between borrowers and lenders is large are classified as hard information environments.

[29] A plant may be located far away in terms of geographical distance, but monitoring may be easier when there are direct flights between the cities where the headquarters and plants are located.

and the financial crisis reminded us of these costs. The growing academic literature has documented the causes and implications of these problems.

As securitization increased the distance between the originator and the ultimate investor that bears the default risk increased, and the performance of credit models deteriorated. The interest rate on loans became a worse predictor of default. Statistical models estimated in a low securitization period broke down during the subsequent period of high securitization (Rajan, Seru, and Vig 2015). The literature has documented a number of explanations for the declining accuracy. First, the history upon which the models were built was short. It is a common theme that new default models are built, often during times of calm in the credit markets, only to fail when the calm passes (Frame, Gerardi, and Willen 2015). It is not just that the historical record was short. When loans are securitized using only information that can be passed on to investors (i.e., hard information), there is little incentive for the lenders to collect soft information (Rajan, Seru, and Vig 2010; Purnanandam 2011; Saengchote 2013; Furfine 2014).

The problem is not just a loss of (soft) information. The historical data is not useful if the behavior of market participants changes in response to the model's introduction (Lucas 1976). An example is the evolving distribution of FICO scores among loan applicants. The underlying distribution of FICO scores in the population is continuous. However, as securitization grew, the number of low documentation loans just above the 620 cutoff jumps relative to the loans just below 620 (Keys, Mukherjee, Seru, and Vig 2010, see Figures I and II). FICO scores above 620 are much easier to securitize and the authors document that they are more likely to be securitized. Loans with a FICO score just above 620 should be of slightly higher quality than loans with FICO scores just below 620, but these loans actually default at appreciably higher rates. This implies that lenders respond to securitization (the sale of loans) by more carefully screening borrower's credit

risk using soft information when they are more likely to hold the mortgage (Arentsen, Mauer, Rosenlund, Zhang, and Zhao 2015). As we discussed with bond ratings (Section III-D), when the decision rule is transparent, borrowers have an incentive to alter the numeric inputs to the credit model. Borrowers "…whose income is more variable and easier to overstate are more likely to end up in the 620+, low-documentation subprime loan pool" (Keys, Seru, and Vig, 2012). Ben-David (2011) documents another example in the mortgage market. He shows that borrowers (and home sellers) were able to inflate the stated value of the house and thus lower the reported LTV (loan-to-value) ratio.

Both hard and soft information are inputs into the regulation of financial institutions as well. The ability of the regulators to see inside the credit institution will dictate the tools they have to regulate financial institutions and eliminate inconsistencies of internal ratings across banks, which is crucial under Basel II and III (Firestone and Rezende 2016; Plosser and Santos 2016). Just as outside investors cannot observe and use soft information generated inside a lender, regulators may have the same problem. This means regulators will need to pay special attention to regulatory policies that depend excessively on default models based on hard information (Keys, Mukherjee, Seru, and Vig 2009). These models ignore the strategic behavior of lenders when it comes to reporting due to the potential loss of rents through the acquisition of private soft information (Hauswald and Marquez 2006; Giannetti, Liberti, and Sturgess 2017).

The lessons from the financial crisis suggest a number of avenues for exploring the benefits and costs of hard relative to soft information. An obvious lesson is that models based on only hard information are potentially subject to manipulation by market participants when the models are transparent and the benefits of small changes in the inputs are sufficiently inexpensive and valuable. The challenge is seeing this problem the next time. Exploring these issues may help us understand

what kinds of loans (or other financial securities or decisions) and what kind of market environments best restrict this behavior. It also raises the question of when and how experience or memory can help or hurt based on the incentives it creates (Chernenko, Hanson, and Sunderam 2016; Diamond, Hu and Rajan 2018)

C) New Financial Markets: FinTech.

The number of financial firms and markets which are labeled as FinTech has increased significantly. The financial problems which these new firms and markets are trying to solve are not new (e.g., evaluating credit quality, allocating investor's assets, raising equity capital for new firms, and increasing the efficiency of payments, among others), but the application of new technology allows for potentially new solutions as well as old problems. Since many of these business models depend upon substituting numerical data and automated decisions based on hard information for decisions made by individuals, they are often built on the concept of substituting hard information for soft. Thus the ideas we discussed above are relevant in understanding why these new solutions may work and the challenges they will face.

1) Numbers versus text revisited.

Some of the logic which underlies FinTech can be traced to a set of academic papers which expanded the data used by academics. A typical firm's 10K filing can run into hundreds of pages. Its financial statements (e.g., its income statement and balance sheet) take up half a dozen pages at most. However, a large fraction of the vast studies that try to explain the changes in equity values with firm data relied only on these accounting numbers and macroeconomic data. This changed when academics started including textual information in regressions by coding the text into numerical scores.

A very early attempt was Das and Chen (2007). They examined the effect of message board postings on the stock prices of Amazon and Yahoo. Although the algorithm was crude, it showed a potential way to incorporating the vast amount of textual data into empirical research. With the digital availability of text and gains in automated methods of analyzing text, there has been an increase in this kind of research. The next iteration, and arguable the paper that kicked off the revolution, was Tetlock (2007) who "…quantitatively measure(d) the interactions between media and the stock market using daily content from a Wall Street Journal column." As the datasets have grown and finance researchers have become more adept at coding the text into numbers in meaningful ways, the literature has grown significantly (Li 2008; Tetlock 2010; Da, Engelberg, and Gao 2011; Dougal, Engelberg, Garcia, and Parsons 2012; Huang, Zang, and Zheng 2014; Hoberg and Phillips 2016; Gentzkow, Kelley, and Taddy 2017; Gianni, Irvine and Shu 2017). The literature has expanded our understanding of how information reported in the financial disclosures and the media (news stories, opinion columns, internet searches and social media) is impounded into stock prices and financial decisions. In each of these papers, the text is condensed into numerical indexes, which capture relevant information (given the empirical results) but likely capture only a portion of what humans may capture. [30] This is a fundamental challenge of hardening soft information. This extraction may undoubtedly lead to a loss of information.[31] Since

[30] The literature began by simply counting positive and negative words, which proved to be more complicated than one would have initially guessed. The language of finance is not as simple as we think (Longhran and McDonald 2011). For example, the sentence "The Dell Company has 100 million shares outstanding" would have been classified as an extremely positive sentence by the early dictionaries, since "company", "share", and "outstanding" are coded as positive words (Engelberg 2008). Hoberg and Phillips (2010) are similar in method but are interested in a very different question. They use text-based analysis of firm's 10-K to measure how similar firms involved in mergers are and thus predict the success of the mergers. Mayew and Venkatachalam (2012) took this idea one step further and examine the information embedded in the voice tone of managers during earning calls.

[31] Loss of information is not only due to the effect of hardening the information. A change in the compensation structure of agents may also affect the use of information. In a controlled experiment, Agarwal and Ben-David (2018) study the impact that changing the incentive structure of loan officers to prospect new applications has on the volume of approved loans and default rates. They find that after the change, loan officers start relying more on favorable hard information and ignoring unfavorable soft information. The results highlight how incentives dictate not just what information is collected but what role it plays in the decision. Another form of loss of information is due to the

the nature of information is not an exogenously fixed quantity, using text analysis or coding soft information into numeric scores, may change or degrade the information. Whether the ability to harden soft information is useful in predicting default models or explaining stock returns, for example, is an empirical question.

For example, in Appendix A, we show the assessment criteria of subjective information measures from an international bank with operations around the globe. This rating is subjective in nature since the criteria are completed by loan officers for each of their clients. This soft rating is part of a final credit rating composed of this soft input plus a hard information score. It is unclear whether this information enhances the information environment of the international bank even compared with a bank's internal rating. Is this information a substitute or a complement of the bank hard information?

2) FinTech lending.

The use of technology (hard information, automated data collection, and automated decision-making) is not new in lending (Einav, Jenkins, and Levin, 2013), but a number of new lending models have arisen under the label FinTech. Peer-to-peer lending (P2P) is one example. In its earliest incarnation it involved individuals lending to other individuals through online platforms. The electronic interface combined with borrowers and lenders who often had no prior relationship, suggested that credit decisions would depend solely on hard information. Like the research on equity markets discussed in the prior section, lenders expanded the information upon which their lending decisions were made beyond the traditional credit metrics.

portability of soft information. For example, Drexler and Schoar (2016) show that when loan officers leave they generate a cost to the bank since it impacts the borrower-lender relationship. As the departing loan officers has no incentives to voluntarily transfer the soft information, borrowers are less likely to receive new loans from the bank in their absence.

Several papers have examined the expanded information used by P2P lenders and what value it has. After controlling for credit score, credit history, income, employment status, and homeownership, personal characteristics, such a physical attractiveness, increase the likelihood of getting a loan and reduce the interest rate (Ravina 2018). Borrower narratives which claim the borrower is trustworthy and successful increase the probability of the loan being funded as well as lowering the interest rate (Herzenstein, Sonenshein and Dholakia, 2011). This is despite the fact that these factors (physical appearance and claims of trustworthiness) are not correlated with lower probabilities of default. Other researchers present a more positive picture of P2P lender's abilities. Iyer, Khwaja, Luttmer, and Shue (2015) find the interest rate set by the market is a better predictor of default, then the borrower's credit score. Lenders are able to use additional information such as the borrower's appearance (picture) or their description of the reasons for borrowing. Duarte, Siegel, and Young (2012) find that borrowers who based on their pictures are perceived to be more trustworthy are more likely to get funded and at a lower interest rate, but unlike the prior work they have lower default rates.

Not all participants in online lending platforms are strangers. In cases where participants are connected off line, these relationships may influence credit outcomes. As with relationship lending, these connection can either provide or signal information, as well as create social pressure to repay loans (Uzzi 1999; Guiso, Sapienza, and Zingales 2013). Having more friends on the lending platform increase the probability a loan is funded, lowers the interest rate, and is associated with lower default rates (Lin, Prabhala, and Viswanathan, 2013). When a borrower's friend defaults, the borrower is more likely to default (Lu, Gu, Ye, and Sheng 2012). Groups of lenders experience lower default rates when lending to an individual personally connected to a group member (Everett 2015). Connections or friendships should matter, but only if they are a credible

and sufficiently costly signal. Returns on loans are higher if an investor who is a friend endorses the loan and bids on it, but returns are lower if friends endorses but does not bid on the loan (Freedman and Jin 2011). The importance of context is one of the distinctions between hard and soft information that we discussed in the first section. Knowing how good a friend you consider this person to be is open to interpretation; understanding you invested your money in the project is not.

These results describe a market that combines characteristics of both relationship and arms-length lending but intermediated through an online platform. It is also a market that is evolving from individuals lending to individuals to an increased reliance on institutional funding of the loans. Eighty percent of the capital on the two major US platforms is now provided by institutional investors (e.g., pension funds and hedge funds) (Ford 2014; Morse 2015), who are arguably more sophisticated (Vallee and Zeng 2018). How the market develops will determine the potential cost reductions and what information can be credibly communicated. What is clear is the set of data upon which lenders make credit decisions has expanded. There are a few possible directions in which these market may evolve. They can depend upon social connections and soft information, but at the cost of limiting the size of the market and the cost savings of scale and automation. The long-term predictability of some of the expanded source of information is not clear. As we discussed in Sections III-D and V-B, borrowers have an incentive to alter the inputs to the credit decision when the benefits exceed the costs. These results raise the question of why borrowers do not learn how to alter their pictures, the search engine they use, and the narrative of why they need to borrow to generate more favorable terms (Duarte, Siegel, and Lang, 2012; Morse 2015; Berg, Burg, Gombovic, and Puri 2018). This may happen in which case these disruptive technologies will be disrupted by a familiar problem (Uzzi 2016).

Alternatively, the market can evolve into using only hard information, numbers which are verifiable and whose interpretation does not depend upon the context not encoded in the data. This is the direction taken by the FinTech mortgage lenders. By replacing soft information completely with hard information, these lenders have removed data collection or real time decision made by humans. The advantages of hard information are apparent in this business model. The loan processing is faster and less expensive due to the automation. This may explain the growth in their market share from 2% in 2010 to 8% in 2016 (Fuster, Plosser, Schnabl, and Vickery 2018). Interestingly, defaults are lower than would be predicted from observable credit metrics (FICO scores and loan-to-value ratios), because lenders can more accurately compare submitted financials to databases and thus prevent fraud and mistakes (Buchak, Matvos, Piskorski, Seru, 2017; Fuster, Plosser, Schnabl, and Vickery 2018).

3) FinTech in equity markets.

The next two applications of new technology to finance are in their infancy and both depend upon the ability to replace transactions that depend upon soft information and relationships with hard information and more impersonal automation: raising equity and investing one's wealth. Public equities are traded among strangers and the investment decisions are based heavily on hard information. The question is whether an analogous market can be developed for the private equity of startups. Equity is more opaque than debt, and the equity of startups even more so. Since there is little hard information (or history) available on these firms most investors rely on personal connections, face to face meetings, and thus soft information. Traditional sources of capital to these firms, venture capitalist and angel investors are therefore geographically close and often intimately involved with the firms before and after funding (Lerner 1995; Sohl 2010; Wong 2002).

Although crowdfunding has drawn a lot of interest and raised a significant amount of capital, its future is unclear.[32] Only a small fraction of the capital raised is in the form of traditional equity (5%), an exchange of capital for a stake in the future earnings and cash flows of the firm. Instead, most of the capital is raised in exchange for a reward or as patronage (Mollick 2014).[33] To the extent crowdfunding draws in small, possibly unsophisticated, investors the opportunity for fraud is nontrivial.[34] The hope is to develop a set of hard information which can capture the economic viability of the potential investments. Given the difficulties we have discussed above in debt markets, which are informationally less sensitive securities, this seems problematic.[35] The efficient market hypothesis argues that an uniformed investors can purchase all public equities and earn the market return. This only works because there are finite set of stocks, which have been vetted by professionals (through the IPO process) and informed investors can buy or sell the stocks and thus push their prices toward the correct value. The logic is different with private firms. There is little to no filtering or vetting of the value of the private firms in this market. Since there is potentially an infinite number of bad investments, market efficiency provides little protection. This raises the question of what role is there for regulation versus a market solution (Agarwal, Gupta, and Israelsen 2016). We already see the emergence of intermediaries that can potentially take on the task of collecting and processing the soft information necessary to evaluate such investments (Catalini and Hui 2018). It is an open research question whether and how the incentive problems documented in the lending market will reoccur in this market.

[32] Mollick (2014) defines crowdfunding as "…the efforts by entrepreneurial individuals and groups… to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the internet, without standard financial intermediaries."

[33] Participants contribute capital in exchange for a product or so that they may participate in supporting an event or creative endeavor. The first is a form of trade credit (pre-paying for a product) and in most examples is more akin to market research then equity funding, since the existence and quality of the product are uncertain.

[34] "...crowdfunding could become an efficient, online means for defrauding the investing public." (Newman 2011).

[35] Investors "…rely on highly visible (but imperfect) proxies for quality such as accumulated capital, formal education, affiliation with top accelerator programs, advisors, etc." (Catalini and Hui 2018).

Another question that deserves attention is how the automation and the elimination of subjective decision-making in choosing the "right" investment by crowdfunding investors affects other market participants. The market for small business lending is partially split into those lenders that focus on soft versus hard information. If equity crowd-funding continues to grow, this may affect the strategy or returns of other angel investors and venture capitalists, who must compete with new more automated selection methods in the early stages and choose how to select and invest among the firms that successfully navigate this process.

4) Robo advising.

When personal investing expanded beyond the wealthy, many retail investors relied on brokers for advice and information. This model was similar to the relationship lending model of banking, relying on soft information and long-term relationships. The rise of mutual funds and discount brokers not only brought in an expanded set of investors but also provided direct access to investing not directly intermediated by brokers (Nocera 2013). Fintech is beginning to make inroads into the retail investment advising business, changing the information available to investors, how it is delivered, and how it is converted into investing decisions. Consistent with the themes of this paper, the information is hard or a hardened version of soft information. This allows for automation of the investment process with associated ability to scale the platforms, reduce costs, lose information, and create perverse incentives.

The Fintech data aggregators have the potential to let investors quickly and cheaply sift through a broader wealth of data on the web. Investors that visit a Fintech (aggregator) web site, however, are significantly less likely to visit the original content site (Grennan and Michaely 2018). Since the recommendations are often a concentrated extract of the original information, some information is lost (see Section III-C). Investors actually prefer recommendations opposed to the

underlying data and reasoning (Grennan and Michaely 2018). There is another potential cost. Collecting and processing soft information is costly for the analysts (Bradshaw, Wang, and Zhou, 2017). As we saw with lending, if investors are relying on the recommendations extracted by finance blogs, and thus don't value the original content, the analysts who produce the underlying information (data and reasoning) may have less incentive to invest in and process soft information and this can lead to less accurate and more biased forecasts (Grennan and Michaely 2018).

Robo-advising takes the next step and provides investment advice with no direct human interaction. Thus, by construction, it relies on hard information and automated decision rules. The information upon which retail investment decisions historically are based is often the hard information of financial reports, past returns, coupled with their reading of analysts' reports, and the advice of brokers who may know them well. When it comes to retail investing, information about the investments (e.g., stocks) is important but so is information about the investors. What is their risk tolerance, level of finance knowledge, and susceptibility to behavioral biases? Investment advising, including robo-advising, needs to address both.[36] Empirical examinations of the effect of robo-advising finds that it can increases diversification of investors, but only for the least diversified investors. It also reduces the "…pervasive behavioral biases such as the disposition effect (being more likely to realize gains than losses) and trend chasing…" although investors appear to trade more often when using these platforms (D'Acunto, Prabhala, and Rossi 2017). The reduction in behavioral biases could improve performance compared to human brokers who are known to exhibit these behavioral biases (Linnainmaa. Melzer and Previtero 2018).[37]

[36] The evidence that human brokers factor their client's characteristics into the investment decision is not reassuring. A retail investors asset allocation depends significantly more on who their broker is (e.g., broker fixed effects) than the investors own characteristics (e.g., risk tolerance, age, financial knowledge, investment horizon, and wealth, see Foerster, S., J. Linnainmaa, B. Melzer, and A. Previtero. 2017.).

[37] Since the algorithms are written by humans, there is also the possibility that the algorithms may embody the same behavioral biases that human advisors have (O'Neil 2016, D'Acunto, Prabhala, and Rossi 2017) as well as the biases of those who design the algorithms or which may be inherent in the data (O'Neil 2016).

VI) Concluding Remarks.

Production and processing of information lies at the heart of the theory of the firm as well as the study of financial markets and institutions. Information is the raw material of all financial decisions. In this paper, we defined the characteristics of what we call soft and hard information with the objective of creating a guide to researchers. The objective is not to define an absolute classification of information into one of two categories, hard and soft, but instead to describe the fundamental characteristics of what we, and the literature, means by soft and hard information and thus frame the discussion of information type. Not only is the classification not discrete, but soft information can, at least partially, be transformed into hard information (the hardening of information). As financial institutions and markets are designed or have evolved, they use hard or soft information in their production processes. As with all of economics, this choice will be driven by the relative advantages of soft versus hard information. The list of the relative benefits and our understanding of them has expanded since we first considered these issues, and may become more elaborate as researchers continue to explore the role each plays in financial decisions.

The terms soft and hard information arose initially in the banking literature (e.g. the lending relationship literature), although the ideas trace back to work on the theory of the firm and the contracting and organizational economics literature. Implicit in this discussion was that information was transmitted from one party (the data collector or agent) to another party (the decision maker or principal). This transmission was costly and imperfect for a number of reasons. The literature on information type tried to define these characteristics and measure their impact on the organization of financial institutions and the structure of markets. As technology allowed greater and cheaper transmission of hard information, markets became geographically more

disperse and some institutions became more hierarchical (relying on hard information) and others less so.

Although banking was initially the core of research on soft and hard information its applications and implications have extended far beyond banking. Researchers have applied these concepts to a variety of markets including: public and private equity, real estate, municipal bonds, and the allocation of capital within the firm. The value is both in explaining how the use of different types of information influences the design and outcome of these markets. It also helps us understand how they will continue to evolve in the future. This is why we used these concepts to examine the research on financial crisis and the emergence of FinTech. These are two areas in which we expect to see future research. Financial crises are a reoccurring feature of financial markets. Although there are a variety of reasons for each unique crisis, the relative costs of using soft versus hard information and the incentive problems this choice can create seems to be a reoccurring factor. We think the same may be true of what is now called Fintech.

Technology and the growth of online data, has motivated the drive to covert data in to quantitative indexes. The growth of numeric data combined with a reduction in the cost of computing have also driven a growth in automated decision-making. This hardening of information is important in finance, but its application can be seen in contexts far from finance, such as the assessment of teaching quality, college rankings, policing and sentencing, and screening employment applications (O'Neil 2016).

We can also see the implications of information type, not just in our field of study finance, but also in what for many of us is a part of our employment responsibilities: teaching. Some elements of teaching have not changed for a millennium. Professor Peter Norvig and Sebastian Thurn taught one of the early MOOCS on artificial intelligence in the fall of 2011. In commenting

on what they learned from the experiment, Professor Norvig describes the classroom experience. It often involves a professor lecturing on a specified body of knowledge and the students taking notes. Now consider the material and methods we teach as viewed through the lens of hard and soft information. Some of what we teach is mechanics (e.g., how to amortize a loan, how to estimate a CAPM regression, how to calculate present values). Some of what we teach is intuition, theory, understanding empirical results, or nuance. The first set of material has the flavor of hard information; the second soft information. In financial institutions, the decisions that depend on hard information have been automated. Expensive labor has been replaced with cheaper capital. The decisions that depend upon soft information have not been automated. Will what happened to lending, happen to teaching? To what extent will the teaching of mechanics be automated and moved out of the classroom? As with financial decisions, this allows for the scare resource, human time and thought, to be focused on a higher value use: the soft information of teaching.

We expect that the literature on soft information may find more applications in other markets and generalize to a variety of other settings. Exploring how large organizations adapt their internal organization structure in order to mimic small organizations will deepen the understanding of ways of organizing production and clarify how rent- or power-seeking activities generate organizational inefficiencies. We leave this for future work.

References

Agarwal, S., and I. Ben-David. 2018. Loan prospecting and the loss of the soft information. *Journal of Financial Economics* 129:608-628.

Agarwal, S., and R. Hauswald. 2016. Authority and information. Working Paper, National University of Singapore.

Agarwal, S, and R. Hauswald. 2010. Distance and private information in lending. *Review of Financial Studies* 23:2757–2788.

Agarwal, S., S. Gupta, and R. Israelsen. 2016. Public and private information: firm disclosure, sec letters, and the jobs act. Working Paper, Georgetown University.

Aghion, P., and J. Tirole. 1997. Formal and real authority in organizations. *Journal of Political Economy* 105:1–29.

Allen, F. and E. Carletti. 2006. Credit risk transfer and contagion. *Journal of Monetary Economics* 53: 89–111.

Allen, F., and E. Carletti. 2015. The roles of banks in financial systems. In *The Oxford Handbook of Banking* Allen Berger, 292-311, eds. A. Berger, P. Molyneux, and J. Wilson. Oxford: Oxford University Press.

Arentsen, E., D. Mauer, B. Rosenlund, H. Zhang, and F. Zhao. 2015. Subprime mortgage defaults and credit swaps. *Journal of Finance* 70: 689–731.

Baker, George, Robert Gibbons, and Kevin Murphy, 2002, "Relational Contracts and the Theory of the Firm" *Quarterly Journal of Economics* 117, 39–84.

Becker G., and K.M. Murphy. 1992. The division of labor, coordination costs, and knowledge, Quarterly *Journal of Economics* 107:1137–1160.

Ben-David, I. 2011. "Financial constraints and inflated home prices during the real-estate boom. *American Economic Journal: Applied Economics* 3: 55–78.

Benmelech, E., and J. Dlugosz. 2009a. The credit rating crisis. *NBER Macro Annual 2009*: 161–207.

Benmelech, E., and J. Dlugosz. 2009b. "The alchemy of CDO credit ratings. *Journal of Monetary Economics* 56: 617–634.

Berg, T., V. Burg, A. Gombovic, and M. Puri. 2018. "On the rise of FinTechs – Credit scoring using digital footprints. Working paper, NBER.

Berg, T., M. Puri, and J. Rocholl. 2016. Loan officer incentives, internal ratings, and default rates. Working Paper, Duke University.

Berger, A. and G. F. Udell. 1995. Relationship lending and lines of credit in small firm finance. *Journal of Business* 68: 351–381.

Berger, A. and G. F. Udell. 2006. A more complete conceptual framework for SME finance. *Journal of Banking and Finance* 30:2945–2966.

Berger, A., N. Miller, M. Petersen, R. G. Rajan, and J. Stein. 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics* 76:237–269.

Berger, A., W. Goulding, and T. Rice. 2014. Do small businesses still prefer community banks? *Journal of Banking and Finance* 44:264–278.

Berger, A., and L. Black. 2011. Bank size, lending technologies, and small business finance. *Journal of Banking and Finance* 35:724–735.

Berger, A. 2015. Small business lending by banks: lending technologies and the effect of banking industry consolidation and technological change. In *The Oxford Handbook of Banking* eds. A. Berger, P. Molyneux, and J. Wilson. Oxford: Oxford University Press.

Berger, A., and C. Bouwman. 2016. *Bank liquidity creation and financial crises* Academic Press.

Berger, A., C. Bouwman and D. Kim. 2017. Small bank comparative advantages in alleviating financial constraints and providing liquidity insurance over time. *Review of Financial Studies* 30: 3416–3454.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100:992–1026.

Bradshaw, M., X. Wang, and D. Zhou. 2017. Soft information in the financial press and analysts' recommendation revisions. Working Paper, Boston College.

Brickley, J. A., J. S. Linck, and C. W. Smith Jr. 2003. Boundaries of the firm: Evidence from the banking industry. *Journal of Financial Economics* 70:351–383

Brown, M., M. Schaller, S. Westerfeld, and M. Heusler. 2012. Information or insurance? On the role of loan officer discretion in credit assessment. Working Paper, University of St. Gallen.

Bolton, P., and M. Dewatripont. 1994. The firm as a communication network. *The Quarterly Journal of Economics* 109(4): 809–839.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru. 2017. Fintech, regulatory arbitrage, and the rise of shadow banks. Working Paper, NBER.

Butler, A. 2008. Distance still matters: Evidence from municipal bond underwriting. *Review of Financial Studies* 21:763–784.

Carruthers, B., and B. Cohen. 2001. Predicting failure but failing to predict: A sociology of knowledge of credit rating in post-bellum America. Working Paper, Northwestern University.

Carruthers, B., and B. Cohen. 2009. Credit, classification and cognition: Credit raters in 19th-century America. Working Paper, Northwestern University.

Catalini, C. and X. Hui. 2018. Online syndicates and startup investment. Working Paper, NBER.

Cerqueiro, G., H. Degryse, and S. Ongena. 2011. Rule versus discretion in loan rate setting, *Journal of Financial Intermediation* 20: 503–529.

Chernenko, S., S. Hanson, and A. Sunderam. 2016. Who neglects risk? Investor experience and the credit boom, *Journal of Financial Economics* 122: 248–269.

Coase, R. 1937. The nature of the firm. *Economica* 4:386–405.

Cohen, B. 1998. Marketing trust: Credit reporting and credit rating in the 19th century. Working Paper, Northwestern University.

Cole, R., L. Goldberg, and L. White. 2004. Cookie cutter vs. character: the micro structure of small business lending by large and small banks. *Journal of Financial and Quantitative Analysis* 39:227–251.

Coval, J., and T. J. Moskowitz. 1999. Home bias at home: local equity preference in domestic portfolios. *Journal of Finance* 54:2045–2074.

Coval, J., and T. J. Moskowitz. 2001. The geography of investment: informed trading and asset prices. *Journal of Political Economy* 109:811–841.

Crane, D., and R. Eccles. 1988. *Doing deals: investment banks at work*. Boston, MA: Harvard Business School Press.

Crawford, V., and J. Sobel. 1982. Strategic information transmission. *Econometrica* 50:1431–1451.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66: 1461–1499.

D'Acunto, F., N. Prabhala, and A. Rossi. 2017. The promises and pitfalls of robo-advising. Working paper, University of Maryland.

Das, S., and M. Chen. 2007. Yahoo! for Amazon: sentiment extraction from small talk on the web. *Management Science* 53:1375–1388.

Degryse, H., and S. Ongena. 2005. Distance, lending relationships, and competition. *Journal of Finance* 60:231–266.

Degryse, H., J. M. Liberti, T. Mosk, and S. Ongena. 2013. Is Loan Officer Discretion Advised When Viewing Soft Information? Working Paper, Northwestern University.

DeMarzo, P. 2005. The pooling and tranching of securities: a model of informed intermediation. *Review of Financial Studies* 18:1–35.

Demsetz, R., and P. Strahan. 1997. Diversification, size, and risk at bank holding companies. *Journal of Money, Credit, and Banking* 29: 300–313.

DeYoung, R., D. Glennon, and P. Nigro. 2008. Borrower-lender distance, credit scoring, and loan performance: evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation* 17:113–143.

DeYoung, R., A. Gron, G. Torna, and A. Winton. 2015. Risk overhang and loan portfolio decisions: small business loan supply before and during the financial crisis. *Journal of Finance* 70: 2451-2488.

Dessein, W. 2002. Authority and communication in organizations. *Review of Economic Studies* 69:811–838.

Diamond, D. 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies* 51:393–414.

Diamond, D. 1991. Monitoring and reputation: the choice between bank loans and directly placed debt. *Journal of Political Economy* 99:689–721.

Diamond, D., Y. Hu and R. G. Rajan. 2018. Pledgeability, industry liquidity, and financing cycles. Working Paper, University of Chicago.

Dougal, C., J. Engelberg, D. García, and C. Parsons. 2012. Journalists and the stock market. *Review of Financial Studies* 25:639–679.

Drexler, A., and A. Schoar. 2016. Do relationships matter? Evidence from loan officer turnover. *Management Science* forthcoming.

Duarte, J., S. Siegel, and L. Young. 2012. Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies* 25: 2455–2484.

Ederington, L. and J. Goh. 1998. Bond rating agencies and stock analysts: who knows what when? *Journal of Financial and Quantitative Analysis* 33: 569-585.

Einav, L., M. Jenkins, and J. Levin. 2013. The impact of credit scoring on consumer lending. *RAND Journal of Economics* 44: 249–274.

Everett, C. 2015. Group membership, relationship banking and loan default risk: the case of online social lending, Working paper, Pepperdine University.

Faulkender, M. and M. Petersen. 2006. Does the source of capital affect capital structure? *Review of Financial Studies* 19:45–79.

Firestone, S., and M. Rezende. 2016. Are banks' internal risk parameters consistent? Evidence from syndicated loans. *Journal of Financial Services Research* 50:211–242.

Ford, J. 2014. Wall Street's siren song lures P2P lenders into treacherous seas. *Financial Times*, October 5.

Foerster, S., J. Linnainmaa, B. Melzer, and A. Previtero. 2017. Retail financial advice: does one size fit all? *Journal of Finance* 72:1441–1482.

Frame, S., A. Srinivasan, L. Woosley. 2001. The effect of credit scoring on small-business lending. *Journal of Money, Credit and Banking* 8:813–825.

Frame, S., K. Gerardi, and P. Willen. 2015. The failure of supervisory stress testing: Fannie Mae, Freddie Mac, and OFHEO. Working Paper, Federal Reserve Bank of Atlanta.

Freedman, S., and G. Jin. 2011. Learning by doing with asymmetric information: evidence from Prosper.com, Working Paper, NBER.

Friedman, M. 1990. *Free to Choose*. New York: Harcourt, Inc.

Furfine, C. 2014. Complexity and loan performance: evidence from the securitization of commercial mortgages. *Review of Corporate Finance Studies* 2: 154-187.

Fuster, A., M. Plosser, P. Schnabl, J. Vickery. 2018. The role of technology in mortgage lending. Working Paper, NBER.

Garicano, L. 2000. Hierarchies and the organization of knowledge in production. *Journal of Political Economy* 108:874–904.

Garmaise, M., and T. Moskowitz. 2004. Confronting information asymmetries: evidence from real estate markets. *Review of Financial Studies*, 17:405-37.

Geanakoplos, M., and P. Milgrom. 1991. A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies* 5:205–225.

Gentzkow, M., B. Kelley, and M. Taddy. 2017. Text as data. *Journal of Economic Literature* forthcoming.

Giannetti, M., J. Liberti, and J. Sturgess. 2017. Information sharing and rating manipulation. *Review of Financial Studies* 30: 3269–3304.

Giannini, R., P. Irvine, and T. Shu. 2017. Nonlocal disadvantage: an examination of social media sentiment. *The Review of Asset Pricing Studies* forthcoming.

Giroud, X. 2013. Proximity and investment: evidence for plant-level data. *The Quarterly Journal of Economics* 128:861–915.

Goetzmann, W., S.A. Ravid, and R. Sverdlove. 2013. The pricing of soft and hard information: economic lessons from screenplay sales. *Journal of Cultural Economics* 37: 271–307.

Granja, J., G. Matvos, and A. Seru. 2017. Selling failed banks. *Journal of Finance* 72:1723–1784.

Grennan, J. and R. Michaely. 2018. Fintechs and the market for financial analysis. Working Paper, Duke University.

Grinblatt, M., and M. Keloharju. 2001. Distance, language, and culture bias: the role of investor sophistication. *Journal of Finance* 56:1053–1073.

Gropp, R., C. Gruendl, and A. Guettler. 2012. Does discretion in lending increase bank risk? Borrower self-selection and loan officer capture effects. Working Paper, EBS Business School.

Grossman, S., and O. D. Hart. 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. *Journal of Political Economy* 94:691–719.

Guiso, L., P. Sapienza and L. Zingales. 2013. The determinants of attitudes towards strategic defaults in mortgages. *Journal of Finance* 68:1473–1515.

Hart, O. 1995. *Firms, contracts, and financial structure*. Oxford: Clarendon Press.

Hart, O., and John Moore. 1990. Property rights and the nature of the firm. *Journal of Political Economy* 98:1119–1158.

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

Hertzberg, A., J. Liberti, and D. Paravisini. 2010. Information and incentives inside the firm: evidence from loan officer rotation. *Journal of Finance* 65:795–828.

Herzenstein, M., S. Sonenshein, and U. Dholakia. 2011. Tell me a good story and I may lend you my money: the role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research* 48:138–149.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124: 1423–1465.

Hong, H., J. Kubick, and J. Stein. 2005. Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance* 60: 2801–2824.

Hu, L., X. Huang, and A. Simonov. 2017. Credit score doctor. Working Paper, Michigan State University.

Huang, A., A. Zang, and R. Zheng. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review* 89:2151–2180.

Ijiri, Y. 1975. *Theory of accounting measurement*. Sarasota, FL: American Accounting Association.

Iyer, R., A. Khwaja, E. Luttmer, and K. Shue. 2015. Screening peers softly: inferring the quality of small borrowers. *Management Science* 62:1554–1577.

Jensen, M.C., and W. H. Meckling. 1995. Specific and general knowledge and organizational structure. *Journal of Applied Corporate Finance* 8:4–18.

Jensen, M.C. 2003. Paying people to lie: the truth about the budgeting process. *European Financial Management* 9:379–406.

Karolyi, S. 2017. Personal lending relationships. *Journal of Finance* 73: 5-49.

Keys, B., T. Mukherjee, A. Seru, and V. Vig. 2009. Financial regulation and securitization: evidence from subprime loans. *Journal of Monetary Economics* 56:700–720.

Keys, B., T. Mukherjee, A. Seru, and V. Vig. 2010. Did securitization lead to lax screening? Evidence from subprime loans. *The Quarterly Journal of Economics* 125:307–362.

Keys, B., A. Seru, and V. Vig. 2012. Lender screening and role of securitization: evidence from prime and subprime mortgage markets. *Review of Financial Studies* 25:2071–2108.

Kimball, A. and R. Cantor. 2008. Should Moody's consider differentiating structured finance and corporate ratings? Moody's Global Credit Policy Group.

Kroszner. K., and P. Strahan. 1999. What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *Quarterly Journal of Economics* 114:1437-67.

Landier, A., V. B. Nair, and J. Wulf. 2009. Trade-offs in staying close: corporate decision making and geographic dispersion. *Review of Financial Studies* 22:1119–1148.

Lerner, J. 1995. Venture capitals and the oversight of private firms. *Journal of Finance* 50:301–318.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45:221-247.

Liberti, J. 2017. Initiative, Incentives and Soft Information. *Management Science* forthcoming.

Liberti, J., and A. Mian. 2009. Estimating the effect of hierarchies on information use. *Review of Financial Studies* 22:4057–4090.

Liberti, J., A. Seru, and V. Vig. 2017. Information, credit and organization. Working Paper, Stanford University.

Lin, M., N. Prabhala, S. Viswanathan. 2013. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science* 59:17–35.

Linnainmaa, J., B. Melzer, and A. Previtero. 2018. The misguided beliefs of financial advisors. *Journal of Finance forthcoming.*

Longhran, T. and B. McDonald. 2011. When is a liability not a Liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66: 35–65.

Loutskina, E., and P. E. Strahan. 2011. Informed and uninformed investment in housing: the downside of diversification. *Review of Financial Studies* 24:1447–1480.

Lu, Y., B. Gu, Q. Ye, and Z. Sheng. 2012. Social influence and defaults in peer-to-peer lending networks. *International Conference on Information Systems* ICIS 2012:1797–1813.

Lucas, R. 1976. Econometric policy evaluation: a critique. In *The Phillips Curve and Labor Markets, Carnegie-Rochester Conferences on Public Policy* eds. K. Bruner and A. A. H. Meltzer. Amsterdam: North Holland Press.

MacKinlay, C. 1997. Event studies in economics and finance. *Journal of Economic Literature* 35: 13–39.

Malloy, C. 2005. The geography of equity analysis. *Journal of Finance* 60: 719–755.

Mayer, C., K. Pence and S. Sherlund. 2009. The rise in mortgage defaults. *Journal of Economic Perspectives* 23: 27–50.

Mayew, W., and M. Venkatachalam. 2012. The power of voice: managerial affective states and future firm performance. *Journal of Finance* 67:1–43.

Mian, A. 2006. Distance constraints: the limits of foreign lending in poor economies. *Journal of Finance* 61:1465–1505.

Milgrom, P., and J. Roberts. 1988. An economic approach to influence activities in organizations. *American Journal of Sociology* 94:154–179.

Morris, S., and H. S. Shin. 2002. Social value of public information. *American Economic Review* 92:1521–1534.

Mollick, E. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* 29:1–16.

Morrison, A. and W. Wilhelm. 2007. Investment banking: past, present and future. *Journal of Applied Corporate Finance* 19:8–20.

Morse, A. 2015. Peer-to-peer crowdfunding: information and the potential for disruption in consumer lending. Working Paper, NBER.

Newman, C. 2011. Is crowdfunding the next big thing or an invitation to digital fraud? *Wired Magazine* December 12.

Nocera, J. 2013. *A piece of the action: how the middle class joined the money class*. Simon & Schuster, Reissue Edition.

O'Neil, C., 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Broadway Books.

Paravisini, D., and A. Schoar. 2015. The incentive effects of scores: randomized evidence from credit committees. Working Paper, London School of Economics.

Plosser, M., and J. Santos. 2016. Bank monitoring. Working Paper, Federal Reserve Bank of New York.

Petersen, M., and R. G. Rajan. 1994. The benefits of firm-creditor relationships: evidence from small-business data. *Journal of Finance* 49:3–37.

Petersen, M., and R. G. Rajan. 2002. Does distance still matter? The information revolution in small business lending. *Journal of Finance* 57:2533–2570.

Puri, M., J. Rocholl, and S. Steffen. 2011. Rules versus discretion in bank lending decisions. Working Paper, Duke University.

Purnanandam, A. 2011. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies* 24:1881–1915.

Qian, J., P. E. Strahan, and Z. Yang. 2015. The impact of incentives and communication costs on information production and use: evidence from bank lending. *Journal of Finance* 70:1457–1493.

Radner, R. 1993. The organization of decentralized information processing. *Econometrica* 61:1109–1146.

Rajan, R.G. 1992. Insiders and outsiders: the choice between informed and arm's-length debt. *Journal of. Finance* 47:1367–1400.

Rajan, R. G., and L. Zingales. 1998. Power in a theory of the firm. *The Quarterly Journal of Economics* 113:387–432.

Rajan, U., A. Seru, and V. Vig. 2010, Statistical default models and incentives. *American Economic Review Papers and Proceedings* 100:506–510.

Rajan, U., A. Seru, and V. Vig. 2015. The failure of models that predict failure: distance, incentives and defaults. *Journal of Financial Economics* 155:237–260.

Ramakrishnan, R. T. S., and A. Thakor. 1984. Information reliability and a theory of financial intermediation. *Review of Economic Studies* 51:415–432.

Ranieri, L. 1997. The origins of securitization, sources of growth, and future potential In *A Primer on Securitization* eds. L. Kendall and M. Fishman (31-43). Cambridge, MA: The MIT Press.

Ravina, E. 2018. Love & loans: the effect of beauty and personal characteristics in credit markets. Working Paper, Northwestern University.

Roll, R. 1984. Orange juice and weather. *American Economic Review* 74:861–880.

Roll, R. 1988. R2. *Journal of Finance* 42: 541–566.

Saengchote, K. 2013. Soft information in the subprime mortgage market. Working Paper, Chulalongkorn University.

Schwert, M. 2017. Bank capital and lending relationships. *Journal of Finance* 72: 787–830.

Seru, A. 2014. Firm boundaries matter: evidence from conglomerates and R&D activity. *Journal of Financial Economics* 111:381–405.

Skrastins, J., and V. Vig. 2018. How organizational hierarchy affects information production? *Review of Financial Studies* forthcoming.

Sohl, J. 2010. The early-stage equity market in the USA. *Venture Capital: An International Journal of Entrepreneurial Finance* 1:101–120.

South, A., and Z. Gurwitz. 2015. Default, transition, and recovery. *2014 Annual Global Structured Finance Default Study and Ratings Transitions*. Standard and Poor's Ratings Services, June 22.

Stiglitz, J., and A. Weiss. 1981. Credit rationing in markets with imperfect information. *American Economic Review* 71:393–410.

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

Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62:1139–1168.

Tetlock, P. 2010. Does public information news resolve asymmetric information? *Review of Financial Studies* 23:3520–3557.

Uzzi, B. 1999. Embeddedness in the making of financial capital: how social relations and networks benefit firms seeking financing. *American Sociological Review* 64:481–505.

Uzzi, B., and R. Lancaster. 2003. Relational embeddedness and learning: the case of bank loan managers and their clients. *Management Science* 49:383–399.

Uzzi, B. 2016. Banks may soon use our Facebook friends to assess our creditworthiness. Should we worry? *Kellogg Insight*, June 6.

Vallee, B., and Y. Zeng. 2018. Marketplace lending: a new banking paradigm? Working Paper, Harvard Business School.

Vazza, D., and N. Kraemer. 2016. Default, transition, and recovery. *2015 Annual Global Corporate Default Study and Ratings Transitions*. S&P Global Ratings, May 2.

Wong, A. 2002. Angel Finance: The Other Venture Capital, Working Paper, University of Chicago.

Exhibit A: Assessment Criteria of Subjective Information Measures

BUSINESS RISK ASSESSMENT

1 *Industry*		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Trend in Output		Very Strong Growth	Strong Growth	Growth	Stable	Uncertain / Declining	Declining
Trend in Earnings		Very Strong Growth	Strong Growth	Growth	Stable	Uncertain / Declining	Declining
Cyclicality (Fluctuations)		Very Stable	Very Limited	Small	Moderate	Large	Large & Unpredictable
External Risks		No Risks	Few Risks, Non Cyclical	Few Critical Risks	Variuos Critical Risks	Numerous Critical Risks	Widespread Risks

2 *Competitive Position*		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Market Position		Over 50% / Clearly Dominant	Over 20% / Dominant	Over 10% / Major Player or Strong Niche	Over 5% / Known Player or Established Niche	2 to 3% / Minos Player	Below 2% / Minor Player; Declining Share
Product Line Diversity		Over 3 Growing Lines	Over 3 Lines	At least 2 Growing Lines	At least 2 Stable Lines	Only 1 Stable Line	Only 1 Declining Line
Operating Cost Advantage		Global Leader	Achieves Low Global Costs	Has Lowest Local Costs	Some Cost Advantages	No Cost Advantages	High Cost Producer
Technology Advantage		Global Leader in Many Areas	Global Player in Some Areas	Leader in Local Market	Mostly New; Upgrading Old	Technology Follower	Predominantly Outdated
Key Success Factors		Global Capabilities in All Factors	Global Capabilities in Most Factors	Strong Locally in All Factors	Strong Locally in Some Factors	Strong in Some; Weak in Others	None

3 *Management*		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Professionalism		At all Levels With Extensive Experience	At all Levels in Operations & Management	At all Key Posi- tions in Operations & Management	At Most Key Positions & Most Levels	At Some Key Positions	In Few Positions
Systems and Controls		Meets Highest Global Standards	Meets Highest Local Standards	Very Reliable and Strong	Acceptable	Unreliable	Largely Absent
Financial Disclosure		Meets Highest Global Standards	Always Timely and Accurate	Usually Timely and Accurate	Satisfactory Reporting	Delayed, Inaccu-rate or Incomplete	Unreliable
Ability to Act Decisively		Proven to be Very Strong	Proven to be Strong	Good, but Untested	Good, but Untested	Weak	Hopeless
Risk Management Policies		**RR1-RR2**	**RR3**	**RR4**	**RR5**	**RR6**	**RR7**
Leverage Policy		Extremely Conservative	Very Conservative	Low Tolerance	Some Tolerance	High Tolerance	Unlimited Appetite
Liquidity Policy		Extremely Conser- vative Cushion	Conservative Cushion & Contingency Plan	Some Cushion & Sound Contingency Plan	Maintains Some Cushion	Low Liquidity Acceptable	No Policy
Hedging Policy		All Risks Understood; No Open Positions	Most Risks Understood; No Open Positions	Most Risks Understood; Few Open Positions	Risks Understood but Not Always Covered	Risks Understood but Most Not Covered	No Hedging Policy / Speculative Policy

4 *Access to Capital*		RR1-RR2	RR3	RR4	RR5	RR6	RR7
Capital Markets		Wide Access; Domestic & International	Wide Access; Domestic & International	Primarily Domestic; Some International	Primarily Domes-tic Banking; Some Capital Markets	Limited Largely to Domestic Banking	No access to Capital markets
Banks		Established Re- lationships; Strong Commitments	Established Re- lationships; Strong Commitments	At Least One Bank Strongly Committed	At Least One Bank Strongly Committed	No Bank Strongly Committed or Some Banks Getting Out	Bank Cutting Lines; Some Locked-in

Overall Business Rating
(Do not use +/- in the final Business Rating)