PART I

FINANCE

ONE

Macroeconomics with Financial Frictions: A Survey

Markus K. Brunnermeier, Thomas M. Eisenbach, and Yuliy Sannikov

1.0 Introduction

The ongoing great recession is a stark reminder that financial frictions are a key driver of business-cycle fluctuations. Imbalances can build up during seemingly tranquil times until a trigger leads to large and persistent wealth destructions, potentially spilling over to the real economy. Whereas in normal times, the financial sector can mitigate financial frictions, in crisis times, the financial sector's fragility adds to instability. Adverse feedback loops and liquidity spirals lead to nonlinear effects with the potential of causing a credit crunch. Classic economic writers who experienced the Great Depression firsthand, including Fisher (1933), Keynes (1936), Gurley and Shaw (1955), Minsky (1957), and Kindleberger (1978), emphasized the importance of financing frictions and inherent instability of the financial system. Patinkin (1956) and Tobin (1969) also emphasized the important implications of financial stability for monetary economics.

This chapter surveys the growing literature that studies the macroeconomic implications of financial frictions straddling three branches of economics: macroeconomics, finance, and general equilibrium theory. All share common themes and similar insights, but they are disconnected in the profession partly because they differ in their modeling approaches and in their identification of the root of the instability. The objective of this survey is to lay bare important theoretical mechanisms and highlight the connections and differences across these approaches.

For helpful comments and discussion, we thank Dong Choi, Wei Cui, Peter DeMarzo, Delwin Olivan, Ricardo Reis, the participants of the 2010 Econometric Society World Congress in Shanghai, and the participants of the 2010 macrofinance reading group at Princeton University. The views expressed in this chapter are those of the authors and are not necessarily reflective of views of the Federal Reserve Bank of New York or the Federal Reserve System.

In a frictionless economy, funds are liquid and can flow to the most profitable project or to the person who most values the funds. Differences in productivity, patience, risk aversion, or optimism determine fund flows, but for the aggregate output, only the total capital and labor matter. Productive agents hold most of the productive capital and issue claims to less productive individuals. In other words, in a setting without financial frictions, it is not important whether funds are in the hands of productive or less productive agents; the economy can be studied with a single representative agent in mind. In contrast, with financial frictions, liquidity considerations become important and wealth distribution matters. External funding typically is more expensive than internal funding through retained earnings. Incentive problems dictate that productive agents issue, to a large extent, claims in the form of debt because they ensure that the agent exerts sufficient effort. However, debt claims come with severe drawbacks: An adverse shock wipes out a large fraction of the levered borrower's net worth, limiting his or her risk-bearing capacity in the future.

Hence, a temporary adverse shock is very *persistent* because it may take a long time until productive agents can rebuild their net worth through retained earnings. In addition to persistence, amplification is the second macroeconomic implication covered in this survey. An initial shock is *amplified* if productive agents are forced to fire-sell their capital. Because fire-sales depress the price of capital, the net worth of productive agents suffers even further (i.e., loss spiral). In addition, margins and haircuts might rise (i.e., loan-to-value ratios might fall), forcing productive agents to lower their leverage ratio (i.e., margin spiral). Moreover, a dynamic amplification effect can kick in. The persistence of a temporary shock lowers future asset prices, which in turn feed back to lower contemporaneous asset prices, eroding productive agents' net worth even further and leading to more fire-sales.

The amplification effects can lead to rich volatility dynamics and explain the inherent *instability* of the financial system. Even when the *exogenous risk* is small, *endogenous risk* resulting from interactions in the system can be sizable. Credit risk can be dwarfed by *liquidity risk*. Liquidity is *fragile* because an infinitesimally small shock can lead to a large discontinuous drop in the price level and a evaporation of funding. Similar systemic risk effects can arise in a setting with multiple equilibria, in which simply a sunspot can lead to these large shifts. Secured funding markets are subject to "collateral runs" when collateral values drop and margins rise. Unsecured funding markets are subject to traditional bank runs, or "counterparty runs," when they are unable to roll over their debt.

To understand these destabilizing effects, it is useful to distinguish among three liquidity concepts: technological, market, and funding liquidity. Physical capital can be liquid because either the investment is reversible (i.e., *technological liquidity*) or the capital can be sold off easily with limited price impact (i.e., *market liquidity*). The latter is the case if the asset has low specificity and, hence, a high value in its second best use. The market liquidity of claims on the payoffs generated by capital goods depends on the liquidity of the underlying physical asset, especially for aggregate shocks, but also on the funding structure of the holder of these claims. Assets with high technological or market liquidity lead to a small fire-sale discount; hence, the amplification effects are contained. Instead of getting rid of the asset by either reverting physical capital or fire-selling it, it also can be used as collateral. Funding liquidity is determined primarily by the maturity structure of debt and the sensitivity of margins and haircuts. If the margin can move from 10 to 50 percent overnight, then 40 percent of the loan essentially has a maturity of one day. Because margins depend on the volatility of the collateral assets, all three concepts of liquidity interact. The determining factor for these destabilizing effects is the *liquidity mismatch* – not necessarily the leverage and maturity mismatch – between the technological and market liquidity on the asset side of the balance sheet and the funding liquidity on the liability side.

The ex-post macroeconomic implications of an adverse shock amplified through liquidity spirals also affect the ex-ante *demand for liquid assets*. In anticipation of potential adverse shocks, market participants have the desire to hold claims with high market liquidity or to preserve high funding liquidity. When individuals face funding constraints, simply the desire to smooth consumption makes it optimal for them to hold a "liquidity buffer." This is the case even in a setting without aggregate risk – for example, when individuals face only (uninsurable) idiosyncratic shocks. Holding liquid assets, which can be sold with limited price impact, allows individuals to self-insure against their idiosyncratic shock when they hit their borrowing constraint. As a consequence, assets that pay off in all states, such as a risk-free bond, are very desirable and trade at a (liquidity) premium. In other words, the risk-free rate is very low and liquid assets are "bubbly." Indeed, fiat money is an asset that provides such a liquidity service. It is a store of value despite the fact that it is not a claim on any real cash flow.

In a more general setting with aggregate shocks (on top of idiosyncratic shocks), the desire to hold liquid assets is even stronger, especially when there is an aggregate liquidity mismatch – for example, the specificity of physical capital is very high (i.e., low market liquidity) and capital

investments are irreversible (i.e., low technological liquidity). At times when exogenous risk increases, these forces strengthen and there will be a *flight to quality and liquidity*. With higher volatility, individuals are more likely to hit their borrowing constraints; therefore, they demand more liquid assets for precautionary reasons.

It is important that the positive price distortions for liquid assets lead to a *constrained inefficient* outcome. That is, a social planner who faces the same constraints as the markets can implement a Pareto-superior allocation. The (constrained) market inefficiency is driven by pecuniary externalities and because individuals take prices as given. This is a strong message because it overturns the standard welfare theorems. In certain environments, the issuance of additional government bonds can even lead to a *crowding-in effect* and be welfare enhancing. As (idiosyncratic) uncertainty increases, the welfare-improving effect of higher government debt also increases. Unlike the standard (New-) Keynesian argument, this reasoning does not rely on price stickiness and a zero lower bound on nominal interest rates.

The role of financial institutions is to mitigate some of these financial frictions. For example, banks can insure households or firms against sudden idiosyncratic shocks by diversifying across them. However, by investing in long-term projects with low technological and market liquidity and by issuing short-term debt claims, financial institutions are exposed to a liquidity mismatch. This maturity transformation – better labeled *liquidity transformation* – is one function of financial intermediation, but it results in fragility. Banks are subject to runs especially if they also are exposed to aggregate risk. A second function of financial institutions is to overcome financial frictions because they have a superior monitoring technology. They can ensure that the borrower of funds exerts enough effort such that projects pay off with a high probability and loans can be repaid. A third function of financial intermediation is the pledgeable creation of informationally insensitive – money-like – securities. Informationally insensitive claims, such as debt contracts, have the advantage that their payoff does not depend on information about underlying cash flows. No one finds it worthwhile to collect information; hence, asymmetric information problems, such as the "lemons problem," cannot emerge. Finally, financial institutions also have a central role in making certain that future cash flows are pledgeable. Productive agents often are not able to pledge future cash flows because of renegotiation. Banks can avoid this problem by offering deposit contracts with a sequential-service constraint, thereby being exposed to bank runs. The threat of a bank run lowers the bankers' ex-post bargaining power and

therefore allows them to pledge a larger amount ex-ante. This literature stresses the "virtue of fragility" as an ex-ante commitment device.

Financial intermediaries are key in understanding the interaction between *price stability* and *financial stability*, and monetary economics more generally. By issuing demand deposits, financial institutions create "inside money" while "Outside money" can take the form of specific commodities or fiat money provided by the government. When banks are well capitalized, they can overcome financial frictions and channel funds from less productive to more productive agents. Financial institutions, through their monitoring role, enable productive agents to issue debt and equity claims to less productive agents. Without a financial sector, funds can be transferred only via outside money. Whenever an agent becomes productive, he or she buys capital goods from less productive agents using outside money and vice versa. Because the fund transfers are limited, money becomes very valuable in this case. In contrast, when the financial sector is well capitalized, outside money is not needed and therefore has low value. Now, a negative-productivity shock lowers financial institutions' net worth, impairs their intermediation activity, and – importantly – makes money more valuable absent any monetary intervention. The latter affects banks on the liability side of their balance sheet because the value of the inside money they issued increases. In short, a negative productivity shock affects banks on the asset and liability sides of their balance sheet and leads to a contraction of inside money. The money multiplier collapses and "Fisher deflation" sets in (as the value of money rises). This effect is in sharp contrast to many other monetary models without a financial sector, which predict inflationary pressure after a negative productivity shock. Monetary policy can mitigate these adverse effects by essentially redistributing wealth toward the financial sector. It is not surprising that money always is highlighted when we speak of liquidity and financial frictions.

Models discussed in this survey assume various financing restrictions. Depending on the underlying economic friction, financing constraints can appear in different forms. For example, debt/credit constraints limit the amount of debt financing. Often, the limit is given by the value of the underlying collateral. In contrast, equity constraints limit the extent to which agents can sell off risky claims. For example, when an agent must have "skin in the game," he or she can sell off only a fraction of the risk. In incomplete-markets settings, risk along certain dimensions cannot be sold off at all; therefore, certain risks remain uninsurable. In models with limited participation, certain agents in the economy are excluded from being active in certain markets altogether. Overlapping generation (OLG) models can be

viewed in the same vein because currently living individuals cannot write contracts with yet-unborn individuals.

The literature offers different *microfoundations* for different financing frictions. First, there is the costly state-verification framework according to Townsend (1979). The basic friction is due to asymmetric information about the future payoff of the project. Whereas the debtor learns the true payoff of the project ex-post, the financier does not. Only if the financier pays a monitoring cost does he or she also learn the true payoff. In such an environment, debt is the optimal contract because it minimizes the socially wasteful monitoring costs. As long as the debt is paid off in full, there is no need to verify the true state. Only in the case of default does the financier verify the state. De jure, the financier must pay the costs, but de facto, he passes them on to the borrower by charging a higher interest rate. This makes external funding more expensive; it drives a wedge between external and internal funding costs and explains why large fractions of projects are funded with retained earnings. It is important that the interest rate increases with the borrowed amount because default and costly monitoring becomes more likely. Increasing the borrowing amount might become unattractive at some point, but the amount of borrowing effectively is not limited.

This is in contrast to quantity rationing, as in Stiglitz and Weiss (1981), for noncollateralized credit. In their setting, asymmetric information arises already ex-ante (i.e., before contracting). Total (market-wide) credit is rationed because the lenders cannot increase the interest rate to ensure that markets clear. They face a "lemons problem," as in Akerlof (1970): Increasing the interest rate would worsen the pool of creditors who apply for a loan such that lenders would lose money. Hence, they ration overall lending and charge a lower interest rate. More specifically, in Stiglitz and Weiss (1981), borrowers have more information about the payoff volatility of their project. Due to limited liability, lenders lose from lending to applicants with high-volatility projects and win from those with low volatility. As they increase the interest rate, the low-volatility borrowers stop applying and the pool of applicants worsens.

Hart and Moore (1994) opened the door for models with incomplete contracts. When payments in certain states of the world are not specified exactly, debtors and financiers will try to renegotiate their future obligations to their favor. Anticipating such future behavior makes certain payoff realizations nonpledgeable. In other words, ex-ante funding often is limited and, as a consequence, a "skin-in-the-game constraint" must be imposed. The limited pledgeability goes beyond the market-wide phenomenon in Stiglitz and Weiss (1981), because it also restricts one-on-one

contract arrangements. One way out of limited pledgeability is to change the ex-post bargaining outcome by collateralizing the initial contract. The literature that uses collateral/margin/haircut constraints typically relies on the incomplete contracting approach as its microfoundation. The literature on limited enforcement of contracts similarly falls in this category. Papers such as those of Bulow and Rogoff (1989); Kehoe and Levine (1993); Alvarez and Jermann (2000); and Cooley, Marimon, and Quadrini (2004) – among others – come to mind.

Empirically, there is convincing evidence on the existence and pervasiveness of financial constraints. The empirical macro literature on credit channels distinguishes between a bank lending channel and a balance sheet channel, depending on whether the financial friction is primarily on the side of the financial intermediary or of the borrowing firm or household. Bernanke (1983) studies the lending channel using data from the Great Depression. Slovin, Sushka, and Polonchek (1993) find that borrowers whose main banking relationship was with infamous Continental Illinois that failed in 1984 earned negative abnormal returns before the (unexpected) government bailout and turned positive on the day before and on the announcement date of the bailout. Peek and Rosengren (1997) document that declines in the Japanese stock market led to reductions in the U.S.-lending-market share of U.S. branches of Japanese banks, with those reductions being larger for banks with weaker balance sheets. Similarly, Gan (2007) finds that following the burst of the real estate bubble, Japanese banks with greater real estate exposure had to reduce lending. Gan also documents the real effects of this credit restriction: In her sample, firms' investment and market valuation are negatively associated with their top lender's real estate exposure. This can lead to effects that are economically significant: In the context of the Japanese depression, the lending channel accounts for one fifth of the decline in investment.

The corporate finance literature has tried mostly to reject the neoclassical theory of investment by showing that financing factors affect investment decisions. A first deviation comes from the fact that capital expenditures react positively to exogenous shocks to cash flows. Most notably, Lamont (1997) shows that following a sharp decrease in oil prices, the non-oil division of oil conglomerates cut their investment. Bakke and Whited (2012) use a regression-discontinuity design that exploits the mandatory contributions to defined benefit plans and find that firms with large cash outflows cut down R&D, working capital, and employment. In a small sample, Blanchard, de Silanes, and Shleifer (1994) report that firms' acquisition activity responds to large cash windfalls coming from legal settlements unrelated to

their ongoing lines of business. Another strand of the empirical literature focuses on the collateral value. For example, Benmelech, Garmaise, and Moskowitz (2005) show that commercial-property loans have lower interest rates, larger loan-to-value ratios, and longer maturities and durations if the underlying properties have fewer zoning restrictions. That is, the properties that are more redeployable and hence have higher market liquidity are superior collateral assets.

A good survey must have a clear focus. This survey's focus is on the macroeconomic implications of financial frictions. This also explains its structure: Persistence, amplification, and instability in Section 2.0 are followed by credit quantity constraints through margins in Section 3.0. The demand for liquid assets is analyzed in Section 4.0 and the role of financial intermediation is studied in Section 5.0. Due to its emphasis on liquidity, the role of money as store of value shines through the entire survey. Given the survey's focus, we do not cover many important papers that microfound the various financial constraints mentioned herein. This survey also does not encompass the vast corporate finance literature on how financial frictions shape the capital and maturity structures of firms and financial institutions. Moreover, this survey excludes behavioral models; we do so despite the fact that we think the departure from the rational-expectations paradigm is important. Exceptions are models with unanticipated zero-probability shocks, in which – strictly speaking – agents hold nonrational beliefs. The survey also touches on bubbles, but the focus on rational models limits us, and we omit important models on bubbles and limits to arbitrage. For a more comprehensive literature survey on bubbles, we refer readers to Brunnermeier (2001, 2008). Other books and surveys including Freixas and Rochet (1997); Bhattacharya, Boot, and Thakor (2004); Heathcote, Storesletten, and Violante (2009); Gertler and Kiyotaki (2010); Shin (2010); Veldkamp (2011); and Quadrini (2011) have a related focus and provide a substitute for the missing parts in our survey.

2.0 Persistence, Amplification, and Instability

2.1 Persistence

The initial macroeconomics literature about financial frictions represented by Bernanke and Gertler (1989) and Carlstrom and Fuerst (1997) focused on the fact that a shock – although temporary – can have long-lasting persistent effects. Whereas even in a standard real-business-cycle model, temporary