Economic Quarterly—Volume 98, Number 1—First Quarter 2012—Pages 1–31

Orderly Liquidation Authority as an Alternative to Bankruptcy

Sabrina R. Pellerin and John R. Walter

When a large nonbank financial firm becomes troubled and in danger of default, government policymakers traditionally have had two options: they could 1) allow the firm to enter bankruptcy, or 2) if policymakers believed bankruptcy is likely to produce widespread (system-wide or "systemic") financial difficulties, the government could provide aid (i.e., a bailout) to forestall failure. In 2010, a third option was made available by the Orderly Liquidation Authority (OLA) provisions, contained in the Wall Street Reform and Consumer Protection Act (the "Dodd-Frank Act"). This legislation authorizes the Federal Deposit Insurance Corporation (FDIC) to pursue an agency-administered wind down for certain troubled financial firms. The OLA provisions are modeled, in part, after the process long followed by the FDIC for handling troubled banks.

The OLA provisions are a reaction to policymakers' and legislators' dissatisfaction with the two options previously available for handling failing nonbanks. For example, Ben Bernanke, chairman of the Board of Governors of the Federal Reserve System, argued, in 2009 testimony before the House Committee on Financial Services, that bankruptcy was not an effective option for certain failing financial firms (Bernanke 2009):

> In most cases, the federal bankruptcy laws provide an appropriate framework for the resolution of nonbank financial institutions. However, the bankruptcy code does not sufficiently protect the public's strong interest in ensuring the orderly resolution of a nonbank financial firm

The authors would like to thank Kartik Athreya, Keith Goodwin, Michelle Gluck, Trish Nunley, Jonathan Tompkins, Zhu Wang, and John Weinberg for their insightful comments. The views expressed in this article are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mails: sabrina.pellerin@rich.frb.org; john.walter@rich.frb.org.

> whose failure would pose substantial risks to the financial system and to the economy. Indeed, after Lehman Brothers and AIG's experiences, there is little doubt that we need a third option between the choices of bankruptcy and bailout for such firms.

In a 2010 speech, Chairman Bernanke expanded on his testimony and noted two goals for this "third option," or "orderly resolution" authority (Bernanke 2010):

> The government instead must have the tools to resolve a failing firm in a manner that preserves market discipline—by ensuring that shareholders and creditors incur losses and that culpable managers are replaced—while at the same time cushioning the broader financial system from the possibly destabilizing effects of the firm's collapse.

Legislators focused on these two goals in the language of the Dodd-Frank Act itself when explaining the purposes of the OLA provisions (or the OLA "title"):

> It is the purpose of this title to provide the necessary authority to liquidate failing financial companies that pose a significant risk to the financial stability of the United States in a manner that mitigates such risk and minimizes moral hazard.

In this article we review the features of bankruptcy and the OLA. We identify some problem areas when large nonbank financial firm failures are resolved through bankruptcy. We then describe two important features of the OLA that are meant to improve on bankruptcy as a means of handling these types of failures, and discuss how they attempt to achieve the goals of mitigating risk to financial stability while also minimizing moral hazard—goals that are not easily achieved simultaneously.

1. FAILURE RESOLUTION

Goals of any Failure Resolution Regime

Any resolution regime, whether bankruptcy, bailout, or OLA, must address two fundamental problems that arise when a firm faces financial troubles and becomes unable to repay creditors. These three regimes each take different approaches to solving these problems, and these differing approaches are at the core of each regime. The first problem (detailed below) is preserving "asset complementarities" and "going-concern value" in the face of detrimental creditor incentives to rush in and grab the firm's assets immediately upon a firm's default. Resolution methods must take these incentives into account and prevent the detrimental actions. The second problem is determining whether to "liquidate" or "reorganize" the troubled firm. Beyond addressing these two

problems, an additional concern arises when the troubled firm is a large financial firm or one with many interconnections with other financial firms: What so called *systemic effects* might the liquidation or reorganization have? Will there be a significant negative effect on other financial firms or on the macro economy in response to actions taken to resolve the troubled firm? As noted in the introduction, policymakers are likely to have a strong interest in any systemic effects when deciding on the appropriate resolution method.

Preserving Complementarities and Going-Concern Value

Following a firm's default on a debt, creditors are likely to rush to seize, and separately sell, assets that, if sold together with other assets, could produce a higher sale price (assets that are "complementary"). For example, one can imagine that with numerous creditors vying for a manufacturer's assets, individual components of an assembly line might be sold off separately, when, if sold as a complete assembly line, these components would be of greater value and produce a higher price. Therefore, this incentive can reduce the total amount that creditors, as a group, receive and can also undercut productivity and economic efficiency. Creditors who manage to be the first to seize assets are likely to recover a higher proportion of their debts than creditors who are slower to react. As a result, creditors have a strong individual incentive to move quickly to undertake such seizures. Preserving complementarities can be important whether the firm is liquidated or is preserved via a reorganization process.

If creditors are allowed to rush in and seize assets, they are also likely to grab those assets that are fundamental to the firm's continued operations, so called "going-concern assets." Such assets might include, for example, necessary operating equipment for a manufacturing firm, or buildings for a financial firm. For a firm that is going to be closed and liquidated, protecting going-concern assets is unimportant, but for firms that might be successful if reorganized, creditors will be made better off, as a group, if their removal is prevented. Indeed, if creditors are allowed to seize going-concern assets, a troubled firm that might otherwise become quite productive in reorganization could be doomed to fail by the asset seizures.

In bankruptcy, the automatic stay (discussed in detail below) prevents immediate asset seizures, and creates a court-overseen process for allocating

assets in a way that preserves complementarities and going-concern value.[1,2] The OLA process also involves a stay, but grants the FDIC this preservation role. Bailouts, by (typically) preventing the troubled firm's default on debts, remove the ability of creditors to seize the troubled firm's assets.[3]

Determining Whether to Liquidate or Reorganize

When a firm becomes unable to meets its debt payments, one of two outcomes are possible. First, as already mentioned, the firm might be closed and its assets liquidated. Alternatively, if the firm can be returned to profitability by restructuring (typically reducing) its debts, then, in many cases, it should be reorganized, allowing it to continue operating after a debt restructuring process. If the firm is unlikely to return to profitability, even with a lowered debt burden, because the firm's assets are unlikely to produce a market rate of return, then the firm should be liquidated: The firm should be shut down and its assets sold to the highest bidders. In this case, liquidation will distribute assets to firms that can make more productive use of them, enhancing economic

---

[1] According to Boul (2006): "Traditionally, the automatic stay has served to 'prevent dismemberment of the [bankruptcy] estate and insure its orderly distribution.' *SEC v. First Financial Group*, 645 F.2d 429, 439 (5th Cir.1981), *citing* S. Rep. No. 95-989, 95th Cong., 2d Sess. 50 (1978); H.R.Rep. No. 95-595, 95th Cong., 2d Sess. 341 (1977), U.S.Code Cong. & Admin. News 1978, pp. 5787, 5836, 5963, 6297, 6298. In that capacity, the automatic stay serves the interests of both the debtor and the creditors of the bankruptcy estate. For the debtor, it provides a 'breathing spell' by 'stopping all collection efforts, all harassment, and all foreclosure actions.' S. Rep. No 95-989, 95th Cong., 2d Sess. 54-55 (1978); H.R. Rep. No 95-595, 95th Cong., 1st Sess. 340 (1977), U.S.Code Cong. & Admin. News 1978, pp. 5787, 5840, 5841, 5963, 6296, 6297. However, the stay also serves the interest of creditors, insofar as it 'eliminate[s] the impetus for a race of diligence by fast-acting creditors.' *SEC v. First Financial Group*, at 439. The stay ensures that assets are distributed according to the order of priorities established by Congress. *Id. at 341*."

[2] Note that if the troubled firm had only one creditor, there would be no need for bankruptcy since that one creditor would always take actions that maximize complementarities and going-concern value. Only in the case where there are many creditors, who, because of their large number, cannot easily coordinate with one another, is bankruptcy necessary.

[3] One might imagine that an ideal solution—when a firm has suffered losses such that its capital level is low and default seems likely, but it could be profitable with a lower debt load—one that requires no intervention by bankruptcy courts or government agencies, is for the firm to gather new funding by issuing new equity shares. The new funding could be used to purchase new, profitable assets that will increase revenues available to service debt (lowering the ratio of debt to assets) and reduce significantly the chance of default. This course may be impossible, however, because of the so-called "debt overhang problem" and, as a result, bankruptcy and the reorganization of debt may be the only course available. Because of the overhang problem, existing equityholders will not vote in favor of a new equity issuance. They will not do so, at least in many cases, because most or all of the benefit flows to the debtholders by improving the market value of their debt, and the existing equityholders will suffer dilution because future earnings must be shared with the new equityholders (Duffie 2011, 43–4). The likelihood that new issues of equity might offer a solution is further reduced by an "adverse selection problem." Weak firms issuing new equity, and especially those firms whose assets are opaque, i.e., financial firms, will have to offer to sell shares at a very low price, because equity investors are likely to conclude, based on the fact that the firm wishes to issue new shares, that the firm is in exceptionally poor health (even worse health than it really is). As a result, existing shareholders will suffer a great deal of dilution and vote against new issues.

productivity and efficiency. Any resolution regime is faced with a decision between liquidation and resolution, and, ideally, will choose the one that produces the most economically efficient outcome.

Addressing Systemic Risk[4] *and Moral Hazard*

When faced with the failure of a large financial firm, or one with many connections with other financial firms, government decisionmakers will not only wish to ensure that complementarities and any going-concern value are preserved, and that the choice between liquidation or reorganization is optimally made, but they will also care greatly about systemic effects. Simply bailing out the troubled firm will prevent its failure, preserve complementarities and going-concern value, as well as avoid systemic effects. But any bailouts will create a "moral hazard" problem: the view, among investors, that large financial firms are likely to be protected, such that in the future, creditors of such firms will reduce their risk-monitoring efforts and these firms will be willing to undertake an inefficiently large amount of risk-taking. Therefore, any method employed to resolve a large or interconnected financial firm must balance systemic dangers against the danger of excessive risk-taking. Bailouts prevent current systemic problems but are likely to lead to less efficient resource allocation choices in the future. Relying on bankruptcy can avoid future moral hazard because, as discussed later, bankruptcy provides no source of funds for bailouts, but the bankruptcy of a large financial firm carries the risk of heavy current systemic problems. As such, when Congress crafted the OLA, addressing systemic risk was a priority, but so was resolving firms in a manner that does not simultaneously increase moral hazard. The OLA aims to address systemic risks that may otherwise be present when resolving systemically important financial institutions (SIFIs) through bankruptcy, in part, by 1) giving the FDIC broad discretion in how it funds the resolution process and pays out creditors, as well as by 2) changing the way derivatives and repurchase agreements (repos)—known as qualified financial contracts ("QFCs")—are treated.

Overview of Bankruptcy and OLA

When comparing bankruptcy and OLA, understanding their overarching goals is important. The goal of a bankruptcy proceeding is to maximize recoveries for creditors, through liquidation or the rehabilitation of the debtor. The goal of the OLA, on the other hand, is to resolve "failing financial companies that

[4] There is no clear consensus about the definition of "systemic risk" (See Taylor 2010). For purposes of this article, we will define systemic risk as "the risk that the failure of one large institution would cause other institutions to fail or that a market event could broadly affect the financial system rather than just one or a few institutions" (Government Accountability Office 2011).

pose a significant risk to the financial stability of the U.S. in a manner that mitigates such risk and minimizes moral hazard."

Bankruptcy achieves its goals through a court-overseen process that relies largely on the troubled firm's creditors and other investors to decide how best, and most profitably, to resolve the firm's troubles. Funding for a bankruptcy resolution typically comes only from the assets of the troubled company and from any funds that might be provided by private investors. See Table 1 for an outline of the bankruptcy process.

OLA borrows several important ideas from bankruptcy, but moves beyond bankruptcy because of policymakers' dissatisfaction with possible outcomes under bankruptcy. The OLA attempts to capture the firms whose resolution through bankruptcy could be detrimental to the broader financial system. Therefore, the OLA can be differentiated from bankruptcy based on several notable features that are designed specifically with SIFI, or covered financial company (CFC), resolution in mind. See Table 2 for a review of OLA's main features.

During the 2007–2008 financial crisis, an unwillingness to trust large firm failures to bankruptcy often resulted in government assistance to firms popularly described as "too big to fail," such as Bear Stearns and AIG. Yet the grant of government assistance sent strong signals to the market that other, similar firms would receive assistance as well if they were to experience trouble, thereby expanding credit subsidies for certain firms and moral hazard. For example, bond prices for the largest financial institutions remained relatively high during the crisis and prices for Lehman credit default swaps (CDS) may not have accurately reflected default risk (Skeel 2010). In contrast, allowing Lehman to fail can be seen as an attempt to mitigate moral hazard; however, some argue this was done at the cost of creating systemic risk.[5] These objectives are inextricably linked, and focusing on the reduction of one has the likely result of increasing the other. Therefore, the OLA, which charges the FDIC with administering these provisions, was an attempt to address this conflict. How does the FDIC meet this challenge?

[5] The apparent worsening of the 2008 financial crisis following Lehman's entrance into bankruptcy provides, for many observers, an illustrative example of the deleterious effect of resolution by bankruptcy for large financial firms. Yet there is some debate about the conclusions one should draw from the Lehman experience. Some observers maintain that the cascading losses following Lehman's bankruptcy filing were not a result of troubles or anticipated troubles related to the bankruptcy process itself, but were instead the result of a shock to market expectations and therefore to the risk assessments of those who had previously anticipated that Lehman, and firms like Lehman, would certainly be bailed out (see Testimony from Skeel before the Subcommittee on Commercial and Administrative Law, Committee on the Judiciary, U.S. House of Reps., October 22, 2009). Available at http://judiciary.house.gov/hearings/pdf/Skeel091022.pdf.

Table 1 Corporate Bankruptcy

Types of Bankruptcy	
Chapter 7	Chapter 7 bankruptcy (liquidation), the troubled firm is closed down, with the longer-run outcome being the sale of all the company's assets (liquidation) because creditors or management do not believe it can be successfully reorganized. Assets of the troubled firm are assembled by the *bankruptcy trustee* and then sold in a manner that maximizes the sum of the payouts to the creditors. The trustee typically must sell all of the bankrupt firm before distributing funds to creditors [11 U.S.C. 704(a)1].
Chapter 11	Under Chapter 11 bankruptcy (reorganization), the troubled firm's debts are reorganized: debt maturities are lengthened, or interest rates or principal amounts are reduced. Creditors will only agree to a reorganization if they believe that preserving the firm as a *going concern* will produce larger payments than if the firm is liquidated.
Corporate Bankruptcies are Overseen by Federal Courts	
	The operating arm of the bankruptcy courts is the Justice Department's Trustee program, so that most bankruptcies are largely handled by trustees.
Circumstances Under which a Firm Enters Bankruptcy	
Voluntary Bankruptcy	When a firm's management petitions the court to place the firm in bankruptcy because it is unable to pay all its creditors in full. A firm will file for bankruptcy when unpaid creditors will otherwise seize complimentary or going-concern assets.
Involuntary Bankruptcy	When a firm's creditors petition for bankruptcy. Creditors have incentive to seek a firm's bankruptcy when they believe that other creditors might seize complementary or going-concern assets or that the firm might dissipate assets.

Table 1 (Continued) Corporate Bankruptcy

Automatic Stay

Immediately, upon the filing of a bankruptcy petition with the clerk of the bankruptcy court, creditors' are prohibited ("stayed") from attempting to collect on their claims.
The stay allows a government-appointed trustee to ensure that assets of the bankrupt firm are liquidated in a manner that maximizes the total pool of funds available for creditor repayment. As a result, the stay allows the trustee to produce a better result for creditors in aggregate than if creditors were simply acting in their own self interest. The trustee can be thought of as solving a joint action problem. Similarly, the stay is also the means in bankruptcy by which creditors are prevented from seizing going-concern assets.
Qualified financial contract (QFC) holders are typically exempt from the automatic stay: They can retrieve their collateral in the event of bankruptcy.

- Under bankruptcy law a number of financial instruments are QFCs, including repurchase agreements (repos), commodity contracts, forward contracts, swap agreements, and securities contracts.
- Reasons for the QFCs exemption:
  - Observers worry that preventing QFC holders from retrieving their collateral could create systemic financial problems.
  - Some observers believe that QFCs are not complementary with one another or with other assets, and can be removed without undercutting the troubled firm's going-concern value.

Table 1 (Continued) Corporate Bankruptcy

Priority Rules	
In Liquidation	Payouts coming from asset sales are divided among creditors based upon the creditor's location in the priority order, which is established in the Bankruptcy Code.
	Secured creditors are repaid from the assets that secure their debts prior to payments to unsecured creditors.
	A secured creditor will be fully repaid if the value of his security exceeds the amount he is owed. If not, he joins unsecured creditors and must depend on the sale of other assets for repayment.
	Unsecured claimants are paid based on the following priority list (White 1998, 1):
	First to be repaid are those owed any administrative expenses produced by the bankruptcy process.
	Second, claims are given statutory priority, such as taxes owed, rent, and unpaid wages and benefits.
	Third are unsecured creditors' claims, including trade creditors' claims, long-term bondholders, and holders of damage claims against the bankrupt firm.
	Last, equityholders receive any remaining funds.
In Reorganization	Payments to creditors and equityholders will often differ from those that would arise based simply on priority rules, because reorganization payments typically arise from negotiation between creditors and equityholders (White 1998, 8).
	Reorganization negotiations are driven by two rules: 1) each class of creditors and equityholders must consent to the bankruptcy plan adopted in the negotiation, and 2) if the negotiation produces no plan that is acceptable to all classes, then the firm is liquidated and payments are determined by the priority rules listed above.
	Because of the mutual consent requirement, some classes can be expected to receive more than would be expected if the priorities rules were strictly followed. For example, if assets are insufficient to repay all creditors, abiding by the priority rule would mean equityholders could expect to receive nothing. But creditors are likely to allow equityholders to receive payments in exchange for the investors' agreement to a plan that allows reorganization rather than liquidation, because the reorganization preserves some going-concern value for all classes. In other words, an equityholder agreement is achieved by paying them more than they would get if they held up the plan.
Debtor-in-Possesion (DIP) Loans	
	Loans made to a firm in reorganization, post-bankruptcy filing.
	Such loans are often senior to all pre-bankruptcy debts.

When the FDIC is appointed as the receiver of a failing financial firm designated as a CFC, it assumes complete financial and operational control of the institution. The FDIC has the authority to manage, sell, transfer, or merge all the assets of the failing firm, as well as provide the funds needed for an orderly liquidation, giving it broad discretion.[6] The FDIC's guiding principles in carrying out these responsibilities include using its best efforts to maximize returns, minimize losses, and, unique to this regime, mitigate the potential for serious adverse effects to the financial system and minimize moral hazard.[7] Moreover, the language of the OLA forces the FDIC to balance two competing interests. On one hand, it is to pay creditors no more than what they would receive in bankruptcy[8] and ensure that creditors bear losses in order to promote market discipline. On the other hand, it is to minimize adverse effects on financial stability. In bankruptcy, creditors only inject additional funds when the firm seems viable. The FDIC, on the other hand, may find it necessary to prop up a firm or perhaps protect certain creditors, at least for a time, to prevent any potential systemic consequences even though the firm may not be viable. The Dodd-Frank Act granted the FDIC a line of credit from the Treasury to fund these efforts. Because the FDIC has broad discretion over the way in which it balances these competing objectives, market participants may find it difficult to predict which objective might receive more weight in any given failure.

2. KEY FEATURES OF BANKRUPTCY, ITS WEAKNESSES, AND OLA AS AN ALTERNATIVE

In the United States, the failure of a business firm typically results in that firm entering *bankruptcy*, and actions taken by the firm shift from being determined by management to being guided by rules established under federal law, specifically under the U.S. Bankruptcy Code. What are the core features of bankruptcy? What features lead observers to conclude that bankruptcy is not an appropriate way to handle a SIFI whose failure could pose substantial risk to the financial system? What are the alternative resolution arrangements created by Dodd-Frank's OLA provisions?

[6] The OLA gives the FDIC authority to operate the company "with all of the powers of the company's shareholders, directors and officers, and may conduct all aspects of the company's business." Dodd-Frank Act § 210(a)(1)(B).

[7] Dodd-Frank Act § 204(a) and § 210(a)(9)(E).

[8] Dodd-Frank Act § 210(d)(2). Under § 210(d)(4)(A) additional payments (in excess of what would be received in bankruptcy) are authorized only with approval of the Treasury Secretary and only if determined to be necessary or appropriate to minimize losses to the receiver.

Table 2 OLA

Who Qualifies as a "Covered Financial Company" (CFC)?
A "financial company" whose failure would have serious adverse effects on financial stability.
Process for Designating a Firm as a CFC
1. Recommendation by Federal Reserve and either FDIC, Securities and Exchange Commission, or Federal Insurance Office, based on their findings that the following is true for the financial company:
    - It is in default or in danger of default
    - A resolution under the Bankruptcy Code would produce serious adverse consequences
    - There is no viable private-sector alternative
2. Determination made by the Treasury Secretary in consultation with the President
3. Appointment of FDIC as receiver of CFC
The FDIC's Powers and Duties
- They can 1) sell the CFC, or any portion of the assets or liabilities to a third party; 2) establish a temporary bridge financial company to preserve the company's value prior to being sold to a third party; or 3) liquidate the company.
- Use their best efforts to maximize returns, minimize losses, and mitigate the potential for serious adverse effects to the financial system.
- Must ensure unsecured creditors bear losses and ensure the directors and management team responsible for the company's condition are removed.
- Has authority to make additional payments to certain creditors (over what their priority would demand and possibly more than similarly situated creditors) if determined to maximize value or limit losses (excess may be "clawed back"), see below.
FDIC's Access to Funding
- Treasury: FDIC may immediately borrow funds from the Treasury (up to 10 percent of the CFC's pre-resolution book-value assets within first 30 days; 90 percent once fair-value is determined and liquidation and repayment plan is in place and approved by Treasury)
- If funds from disposition of failed firm's assets are insufficient to repay Treasury:
    - Creditors (who were paid more than they would in bankruptcy) would have to return excess funds ("claw backs")
    - Large financial institutions can be assessed

Notes: "Financial Company" includes bank holding companies, nonbank financial firms, and securities broker-dealers. Nonbank financial firms are characterized as firms that are supervised by the Fed (because of SIFI designation) or that derive at least 85 percent of their revenues from activities that are financial in nature.

Key Bankruptcy Feature: The Automatic Stay

The "automatic stay" is a primary component of bankruptcy and one that underlies many of the complaints raised against bankruptcy as a means of handling SIFI failures. The stay works as follows. Immediately upon the filing of a bankruptcy petition with the clerk of the bankruptcy court, creditors are enjoined from attempting to collect on their claims.[9] This feature of bankruptcy allows a government-appointed trustee to ensure that assets of the bankrupt firm are liquidated in a manner that maximizes the total pool of funds available for creditor repayment. Without the stay, as discussed earlier, creditors can be expected to rush in, grab, and then sell the bankrupt firm's assets. In so doing, creditors could destroy asset complementarities. The stay typically lasts for the length of the bankruptcy process, though the courts may grant exceptions.

In a Chapter 7 bankruptcy (liquidation),[10] the type of corporate bankruptcy in which the troubled firm is closed down (liquidated), the court-appointed trustee typically must sell all of the assets of the bankrupt firm before distributing funds to creditors.[11] The goal of the trustee is to sell the assets in a manner that maximizes the sum of payouts to creditors. Achieving this maximization goal can result in a lengthy process, so that creditors' funds may be inaccessible for an extended period. Based on a study of all corporate bankruptcies from two federal bankruptcy court districts between 1995 and 2001, the average liquidation lasts 709 days (Bris, Welch, and Zhu 2006; 1,270). It seems likely that for the largest, most complex financial firms the process will take at least as long as average or perhaps longer.

Compared to liquidation, a corporate Chapter 11 bankruptcy (reorganization) process tends to last longer still, 828 days on average according to Bris, Welch, and Zhu (2006), though in reorganization creditors will often be repaid well before this process ends. In reorganization, the troubled firm's debts are rescheduled or cut—but it continues to operate.[12] A corporation that finds itself unable to repay all creditors in full can seek protection from creditors' claims by petitioning the bankruptcy court to enter reorganization. This protection from creditors, which includes a stay of claims, is important when a firm is being reorganized because the stay prevents creditors from seizing "going-concern" assets (assets that might be necessary to keep the firm running). The stay can mean that, in aggregate, creditors receive more than

[9] 11 U.S.C. § 362

[10] In the remainder of the article, for the sake of simplicity, we will typically replace the phrase Chapter 7 bankruptcy with "liquidation" and the phrase Chapter 11 bankruptcy with "reorganization." We will use the phrase "orderly liquidation" or the acronym OLA when referring to a Dodd-Frank Orderly Liquidation Authority process.

[11] 11 U.S.C. 704(a)1

[12] The airline industry provides many well-known examples of reorganization, in which planes continue to fly and contracts are renegotiated with creditors and employees.

they would if individual creditors had been allowed to seize assets to protect themselves. Because creditors must agree to the troubled firm's proposed reorganization plan—if not, the firm is likely to proceed to a liquidation—firms receiving reorganization treatment are those for which creditors, as a group, believe going-concern value exceeds the value of firm assets if such assets are sold, i.e., if the firm is liquidated (White 1998, 2–3).

While reorganization can last longer than liquidation, payouts to creditors will often be made well before the end of the reorganization process. As part of the reorganization, creditors may agree to lower repayments and some may receive these repayments quickly. Further, additional funding can flow into the troubled firm fairly quickly to help keep it afloat.

A source of funding often available to a firm in reorganization is "debtor-in-possession" (DIP) funding. In reorganization, the troubled corporation, the debtor, continues to operate, or "possess," the troubled entity. Any loans to the troubled corporation are therefore loans to the DIP. Such loans are often senior to all former—prior to the bankruptcy filing—debts of the bankrupt firm. The prospect of being senior to other creditors allows funding to flow as long as creditors can be convinced that the firm is likely to survive and therefore repay.

Key Bankruptcy Feature: Limited Sources of Funding

Repayment of a bankrupt firm's creditors and funds to sustain a firm reorganized under bankruptcy can only derive from two sources: the assets of the troubled firm, and, in the case of reorganization, added (DIP) loans that might flow to the troubled firm. While bankruptcy law and practice do not prohibit government aid to troubled firms, such funding is not typically available. As a result, creditors have an incentive to carefully evaluate the riskiness of any firm prior to providing funding and to monitor its activities once funding has been provided. Such monitoring will tend to ensure that the firm undertakes only those risks with a positive expected return. Yet, the government has often provided aid to troubled firms because of the sluggishness with which creditors are often repaid following failure and because of the apparent difficulty of lining up DIP funding. In some cases this aid has been provided prior to bankruptcy, in others during bankruptcy.[13] Therefore, the

[13] Bear Stearns and AIG provide examples of financial firms that received government aid prior to bankruptcy. In 2009, both General Motors and Chrysler received aid from the federal government during their reorganizations. Earlier cases of government aid include Penn Central Railroad in 1970, Lockheed Aircraft in 1971, and Chrysler in 1980.

monitoring advantage offered by bankruptcy can be diminished by the expectation of government aid for certain (especially large) financial firms.[14]

There is no DIP financing in a liquidation. In liquidation, a "bankruptcy estate" is created, including all of the assets of the bankrupt firm. One of the responsibilities of the trustee is to locate all assets and gather them into the estate. The estate assets are sold by the bankruptcy trustee and the proceeds of the sale provide the funds from which creditors are repaid. Funds from no source beyond the assets of the failed firm are available to the trustee and therefore to the creditors.

In a reorganization proceeding, debts are restructured in a manner such that the firm can continue operating. For example, the creditors of a firm might come together and all agree to reduce the amounts the bankrupt firm owes each of them by 30 percent, and extend the maturity of all debts by two years. As a result, the bankrupt firm faces lower monthly debt payments, payments that it might successfully manage. The creditors will only agree to such a plan if they believe that sustaining the operations of the firm is likely to mean larger payments than if the firm descends into liquidation. The debt restructuring and the mode of future operation is called the "reorganization plan" and is subject to court review and creditor appeal to the bankruptcy court. Typically the current management of the troubled firm operates the reorganized firm. If the firm's liabilities exceed its assets, owners are wiped out and the creditors inherit the decisionmaking rights formerly enjoyed by owners. The debtor can acquire funding for the reorganized firm because it can offer very favorable terms to the lenders who provide DIP funding because the new lenders have a claim that is senior to all other creditors. Thus, lenders will have an incentive to provide DIP funding if they believe that the reorganized firm is likely to be able to repay their loans from future earnings—that the reorganized firm will be profitable.

Weaknesses of Bankruptcy

A Weakness of Bankruptcy for Financial Firms: The Stay Threatens Short-Term Debtholders

While the automatic stay, in liquidation or reorganization, may cause no spread of losses when the creditors of the troubled firm are typically long-term debtholders (who are not counting on quick receipt of their funds), in the

[14] One might argue that there could be times in which government aid is appropriate, for example if credit standards have become inefficiently (or irrationally) strict, as in a financial panic. If market participants believe that government aid will only be forthcoming at such times, and will only provide the amount of funding that private lenders would provide if they had not become irrationally strict, then the expectation of government aid will not diminish private investors' risk-monitoring efforts.

case of a failing financial firm, creditors are likely to include a large contingent of those with very short-term claims. Funds invested in financial firms (such as investment banks) often have maturities of one or a few days. Creditors with such short maturity claims are likely to be dependent on the immediate access to their funds in order to pay their own creditors. If funds are tied up for an extended period, as assets are gathered and sold in a liquidation process or as a reorganization agreement is negotiated, the bankrupt firm's creditors may find themselves unable to make payments to their own creditors. As a result, the bankruptcy of one firm may result in the failure of some of its creditors, especially if some of these creditors are also financial firms with their own very short-term debts to repay. Therefore, while the automatic stay may have significant value in preventing creditors from separating complementary assets in liquidation and preserving going-concern value in reorganization, the stay, if it continues more than a very short time, may cause financial distress to spread. The importance of short-term funding, which is often present for nonbank financial firms, may make policymakers unwilling to rely on bankruptcy when such firms become troubled.

A Weakness of Bankruptcy for Financial Firms: Opacity Reduces Availability of DIP Financing

New funding, quickly available, will often be necessary in order for a troubled firm to be successfully reorganized. After all, funds from former sources may have dried up because of the losses these creditors suffered on former loans to the troubled firm. But, financial firms may find it to be relatively difficult, compared to nonfinancial firms, to quickly obtain DIP funding. Such firms often have quite opaque assets: assets that are difficult for outsiders, such as lenders, to value. For example, assets of financial firms often include a heavy concentration of loans to other firms. The value of such loans may depend importantly on information that can be gathered only by performing detailed analyses of the financial condition of the borrowing firms.[15] As a result, DIP loans may be available only after lenders spend a great deal of time reviewing the troubled firm's assets. Further, DIP loans made to financial firms are likely to involve unusually high interest rates to compensate for time spent in asset review and for the potential risk of lending to a firm with highly opaque assets.

[15] Using statistical analysis to measure firm opacity, by comparing the frequency of bond rating disagreements, Morgan (2002, 876) finds that banks and insurance firms are the most opaque of major industry groups. Large nonbank SIFIs are likely to have a portfolio of assets that are fairly similar to bank asset portfolios so can be expected to be similarly opaque. Interestingly, Morgan notes that the industry grouping "Other Finance and Real Estate" seems to be among the least opaque, though, according to Morgan, this is likely because the securities being analyzed for this group are "asset-backed bonds backed by a pool of specific, homogeneous assets 'locked' up in special purpose vehicles. This structure, which reduces the risk of asset substitution, seems to make the securities relatively safe and certain to outsiders" (2002, 877).

The opacity of financial firm assets contributes to the desire to employ some method (i.e., bailouts or OLA) for their resolution instead of bankruptcy.[16]

Key Features of OLA and OLA's Weaknesses

As in bankruptcy, when a troubled financial firm enters the OLA process, creditors—with the exception of holders of QFCs, discussed below—are stayed (prevented) from collecting their debts. The stay lasts the duration of the period in which the financial firm is in the OLA process. During the stay, the FDIC will typically establish a receivership estate into which most assets and liabilities will be placed. Assets placed in the receivership will be sold by the FDIC in the manner that results in the largest returns to creditors—so that the receivership may last, and creditors wait, an extended period while the FDIC lines up buyers. In addition, some of the bankrupt firm's assets and liabilities can be moved into a "bridge entity," a separate company formed by the FDIC, which might be sold off as a whole entity to a private buyer or might even be capitalized by some of the creditors of the bankrupt firm, and continue as a going concern.[17] One purpose of a bridge can be to preserve going-concern value of portions of the troubled firm.[18]

The Dodd-Frank OLA process also abides by a priority schedule similar to the one defined in bankruptcy law (see Table 1 for an overview of bankruptcy priorities). But Dodd-Frank authorizes the FDIC to violate the priority list established in OLA under certain circumstances. Specifically, section 210(d)(4) of the Dodd-Frank Act permits the FDIC to pay a creditor more than priority rules might otherwise allow "if the Corporation determines that such payments or credits are necessary or appropriate to minimize losses to the Corporation as receiver from the orderly liquidation of the covered financial company." According to the FDIC's discussion of its proposed rules related to this section of the Dodd-Frank Act, such additional payments may be made if they are necessary to "continue key operations, services, and transactions that will

[16] An alternative to bailouts or OLA that would address the problem of a lack of DIP funding as a result of SIFI opacity is to allow a troubled SIFI to enter reorganization, and permit the government to make DIP loans to the bankrupt firm. The government could quickly provide DIP funds to keep the firm operating but the bankruptcy process could handle all other aspects of the resolution.

[17] See Acting Chairman Martin J. Gruenberg's (2012) presentation before the Federal Reserve Bank of Chicago Bank Structure Conference for a discussion of how a bridge bank might be capitalized and continue operations as a private entity.

[18] Acting FDIC Chairman Gruenberg (2012) discussed the formation of a bridge, and noted its advantages for protecting going-concern (franchise) value: "... the most promising resolution strategy from our point of view will be to place the parent company into receivership and to pass its assets, principally investments in its subsidiaries, to a newly created bridge holding company. This will allow subsidiaries that are equity solvent and contribute to the franchise value of the firm to remain open and avoid the disruption that would likely accompany their closings... In short, we believe that this resolution strategy will preserve the franchise value of the firm and mitigate systemic consequences."

maximize the value of the firm's assets and avoid a disorderly collapse in the marketplace."[19]

Beyond the authority to, in some cases, make greater payments to creditors than their priority might allow, the Dodd-Frank Act also provides the FDIC with Treasury funding that might be used to make payments to creditors. The Act provides that the FDIC can borrow, within certain limits, from the Treasury. Immediately upon their appointment as receiver of a firm, the FDIC can borrow 10 percent of the value of the firm's pre-resolution assets. For a large financial firm, this initial amount can be significant. In the Lehman failure, for example, 10 percent of assets would have amounted to $63.9 billion. Once the fair value of the failing firm's assets is determined and a liquidation and repayment plan is in place, the FDIC may borrow an additional 90 percent of the value of the firm's assets (with approval from the Treasury). The Act provides that these funds are to be repaid to the Treasury from the sale of the liquidated firm's assets. But, importantly, the Act also specifies a means of repayment if such assets are not sufficient for repayment, first by attempting to "claw back" any "additional payments" (payments beyond what would have been received in a liquidation) made to creditors, and, if that is insufficient, by taxing all large bank holding companies and other SIFIs (Dodd-Frank Act § 210(o)(1)(A)).[20,21,22] The fact that assets might not be sufficient to repay the Treasury in full, and that the legislation authorizes taxes (on large financial

---

[19] http://edocket.access.gpo.gov/2011/pdf/2011-1379.pdf; 4,211

[20] The Dodd-Frank Act § 210(o) specifies that assessments (taxes) to repay the Treasury are to be imposed on bank holding companies with assets greater or equal to $50 billion and on nonbank financial companies supervised by the Board of Governors of the Federal Reserve (meaning nonbank SIFIs). Assessments are to be sufficient to repay the Treasury within 60 months, with the opportunity for extension if repaying in 60 months would have a "serious adverse effect on the financial system." Assessments are to be graduated based on company size and riskiness. When determining assessment amounts, the FDIC, in consultation with the Financial Stability Oversight Council, should take account of "economic conditions generally affecting financial companies so as to allow assessments to increase during more favorable economic conditions and to decrease during less favorable economic conditions...the risks presented by the financial company [being assessed] to the financial system and the extent to which the financial company has benefitted, or likely would benefit, from the orderly liquidation of a financial company under this title," and any government assessments already imposed on the firm under such government programs as deposit insurance or securities investor protection insurance.

[21] The Dodd-Frank Act § 210(o)(1)(D)(i) prohibits the FDIC from imposing claw backs on creditors who receive "additional payments" if such payments are "necessary to initiate and continue operations essential to implementation of the receivership or any bridge financial company." The FDIC's implementing regulation, at 12 CFR 380.27, seems to imply that a good portion of any additional payments made by the FDIC will be for such essential purposes so will be protected from claw back. Note that if all additional funds could be clawed back, there might be little reason to be concerned about the potential moral hazard problem created by FDIC payments. But, given that the FDIC is likely to be prohibited from imposing claw backs on some significant portion of payment recipients, the moral hazard concern seems to be in play.

[22] Analysts (Acharya et al. 2009, 31–2; Acharya et al. 2011, 10–1) have noted that it would be more appropriate to impose this tax prior to any failure, and base the tax rate on a firm's riskiness. Such a tax would discourage risk-taking. The current tax does not discourage risk-taking, since the failing firm does not pay it. In fact, because it is paid by survivors, it punishes, and therefore discourages, caution.

firms) to repay the Treasury, implies that creditors may be repaid more than the sum of funds generated by asset sales—more than they would have been repaid in liquidation.

It seems likely that Congress intended to provide the FDIC with a good bit of discretion to bypass strict priority as well as discretion over whether to borrow Treasury funds in order to mitigate systemic risk. For example, given the FDIC's ability to pay some creditors more than they would receive in bankruptcy, these creditors may be less likely to pass on losses to other firms, lowering the risk of a systemic problem.

One might argue that legislators' intention for providing the FDIC with the authority to borrow from the Treasury was simply to allow the FDIC the ability to move quicker than bankruptcy courts. By providing an immediate source of funds, the FDIC could gather funds, which it could then use to make payments equivalent to what would be paid in bankruptcy. In this way creditors would not be denied access to their funds for months or years (as in liquidation), and the FDIC could slowly sell the assets of the failing firm such that fire sales are avoided. Under such an arrangement, legislators could have required the FDIC to immediately estimate the value of the failing firm's assets (similar to the type of analysis currently performed by the FDIC when it determines—and announces in a press release—the cost to the FDIC of a bank's failure), and then limit itself to paying creditors no more than their pro-rata share (given priorities) of this estimated amount. Yet, Congress did not choose this course, i.e., it did not require the FDIC to limit the sum of its payments to be no more than the estimated value of the failing firm's assets. Instead it left the FDIC to determine payments to creditors and authorized taxes on large financial firms if payments exceed the liquidation value of assets. Therefore, it seems clear that Congress intended for some creditors of a failing firm to receive larger payments than bankruptcy allowed, as a means of mitigating systemic risk.

Investors certainly realize that the OLA provisions provide the FDIC with the authority to make larger-than-bankruptcy payments to creditors. As a result, they will tend to under price risk-taking by nonbank firms that might get OLA treatment and such firms will engage in more risk-taking than if they did not enjoy the potential benefits of receiving government aid.[23] Congress was aware that larger payments would have this moral-hazard-exacerbating impact on firm risk-taking and took steps to mitigate the impact in the OLA provisions of the Dodd-Frank Act. Broadly, the legislation requires that the FDIC attempt to liquidate SIFIs "in a manner that ... minimizes moral

[23] Some authors, such as Jackson (2011), argue that a modified bankruptcy procedure can address this excessive risk-taking weakness and better resolve SIFIs. According to them, a system of established rules, judicial oversight, and full public disclosure has a better chance of both reducing bailouts and making the costs of them known than does a non-bankruptcy resolution authority.

hazard."[24] More specifically, the law calls on the FDIC to ensure that any member of the management or the board of directors of the failed firm who is deemed responsible for the failure is fired. Similarly, the OLA provisions require the FDIC to "ensure that the shareholders of a covered financial company do not receive payment until after all other claims and the Fund are fully paid and ensure that unsecured creditors bear losses..."[25,26] The provisions requiring the removal of management and directors are likely to encourage these corporate leaders to limit risk-taking. However, the OLA contains provisions for certain creditors to receive better treatment than they might in bankruptcy, even if some creditors suffer losses, so that creditor oversight is likely diminished by the prospect of OLA treatment.

Dealing With Systemic Risk in Failure Resolution: Exceptions to the Automatic Stay

The class of financial contracts, which are exempt from the automatic stay, are commonly referred to as "qualified financial contracts" (QFCs).[27] Therefore, investors who are holding QFCs have the ability to immediately terminate and net-out their contracts or liquidate the collateral on their claims once a party has defaulted or filed for bankruptcy. Today, under bankruptcy law, a number of financial instruments are QFCs, including repos, commodity contracts, forward contracts, swap agreements, and securities contracts.[28] The treatment of QFCs in bankruptcy (and under OLA provisions) has been the focus of a great deal of public debate.

A possible explanation for exempting QFCs is that the collateral that typically backs QFCs is not directly tied to the defaulting firm's going concern value. A primary objective of the automatic stay in bankruptcy is to prevent

---

[24] Dodd-Frank Act § 204(a)

[25] Dodd-Frank Act § 206(1-5)

[26] The Dodd-Frank Act includes other provisions intended to minimize moral hazard including 1) a requirement that SIFIs create resolution plans ("living wills") to increase the likelihood that they would be resolved through bankruptcy [Dodd-Frank Act § 165(d)]; and 2) a requirement that the FDIC have a plan in place, before borrowing greater than 10 percent of the failing firm's asset, for repaying the Treasury [Dodd-Frank Act § 210(n)(9)(B)].

[27] In the Bankruptcy Code, contracts exempt from the automatic stay are referred to as "safe harbor contracts." The Federal Depository Institution Act and the Dodd-Frank Act refer to the safe harbor contracts as QFCs. Since safe harbor contracts and QFCs generally refer to the same types of contract, we will use the term "QFC" to refer to both, which is consistent with industry practice.

[28] The types of contracts exempt from the stay are listed in the following sections of the Bankruptcy Code: 11 U.S.C. § 362(b)(6), (b)(7), (b)(17), 546, 556, 559, 560. All terms are defined in 11 U.S.C. § 101 with the exception of a "securities contract," which is defined as "the purchase, sale, or loan of a security, including an option for the purchase or sale of a security, certificate of deposit, or group or index of securities (including any interest therein or based on the value thereof), or any option entered into on a national securities exchange relating to foreign currencies, or the guarantee of any settlement of cash or securities by or to a securities clearing agency" (11 U.S.C. § 741).

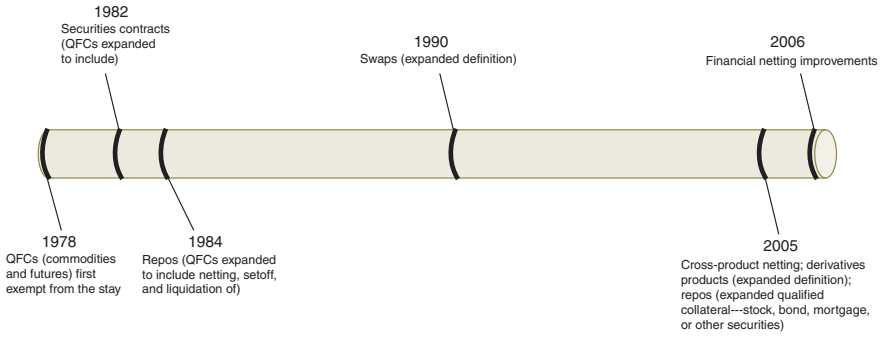
the separation of complementary assets (an important goal of the trustee in liquidation) or to preserve the going-concern value of a firm (typically a goal in reorganization). QFCs can be immediately closed out because the collateral backing them will typically not be complementary to other assets of the firm, nor will QFC collateral be important to the firm's going-concern value. For instance, collateral consisting of highly marketable or cash-like securities (for example Treasury bills or mortgage-backed securities) can be removed from the firm without necessarily undercutting the firm's ability to produce loans or other financial products, since the production of these products depends on such resources as the skill of lending staff, staff contacts with possible borrowers, IT assets, office space and equipment, and funding (liabilities) from which to make loans. However, some argue that the collateral backing certain QFCs can be firm-specific (e.g., a pool of mortgage cash flows used as repo collateral) and therefore not all QFCs should be treated equally (Jackson 2011).

Another possible explanation for exempting QFCs is that the markets in which QFCs trade are special, such that delaying creditor recovery attempts in these markets (by imposing a stay on QFC counterparties) is especially destructive, compared to staying creditors operating in other markets. More specifically, proponents who hold this view seem to be arguing that staying QFCs is more likely to create systemic problems than staying the collection of other debts. This explanation for special treatment—what we will call the "systemic risk" rationale—appears to stand out as the argument used by policymakers supporting the expansion of the list of QFCs that took place over several decades through numerous reforms to the Bankruptcy Code. The rationale offered by those supporting the exemption is that in a fast-paced, highly interconnected market, a counterparty to a QFC may need the proceeds from the contract to pay off other debts in a timely manner. If this counterparty is unable to meet other obligations as a result of having its contracts held up in bankruptcy, other firms relying on that counterparty may become exposed and experience financial distress, which could bleed to other counterparties, and so on, causing a ripple effect and possibly "destabilizing" markets (Edwards and Morrison 2005).[29]

Today, the transactions and agreements covered under the definition of a QFC include a wide range of instruments. However, when the automatic stay

[29] In a letter dated September 30, 1998, to Hon. George W. Gekas, Chairman, Subcommittee on Commercial and Administrative Law, Committee on the Judiciary, Robert Rubin, former Treasury Secretary, argued that applying traditional insolvency laws, such as the stay, to QFCs could cause a "possible domino effect that could turn the failure of one market participant into a failure of the market." See www.wilmerhale.com/files/Publication/eacecfbd-0400-4cb1-80a0-cf3a2c3f1637/Presentation/PublicationAttachment/29b1ce6d-1ce1-4544-a3ec-63ecd65d11e1/Bankruptcy%20%20Derivatives%20outline%20-%20_final_.pdf.

Figure 1 History of QFC Exemptions from the Stay



was first created as part of the new Bankruptcy Code in 1978,[30] only commodities and futures contracts were exempt.[31] At the time, these protections were intended to "prevent the insolvency of one commodity firm from spreading to other brokers or clearing agencies and possibly threatening the collapse of the market."[32] In the decades to follow, various reforms to the Bankruptcy Code expanded the types of contracts classified as QFCs, as well as expanding the types of collateral that could be used to back them (see Figure 1 timeline).

Legislation enacted in 2005 and 2006[33] expanded the safe harbor treatment significantly by broadening the definition of a QFC to such an extent that it would capture any newly created derivatives product that may otherwise not be explicitly included.[34] Moreover, the most recent reforms also expanded contractual netting rights to allow for "cross-product netting" of QFCs (Figure 1). Netting occurs when a non-defaulting counterparty of a defaulting bankrupt firm is allowed to offset debts it owes to the defaulting firm against debts owed it by the defaulting firm.[35] Cross-product netting allows contracts

[30] The *stay* existed as a fundamental feature of bankruptcy before 1978. The Bankruptcy Reform Act of 1978, however, created the "automatic stay," which takes effect immediately upon the filing of a bankruptcy petition. Prior to the Bankruptcy Reform Act of 1978, the stay typically took effect only after the grant of an injunction by a court. Such grants were typical, but were often not immediate, and certainly not automatic (Jessup 1995).

[31] U.S.C. §362(b)(6)

[32] See H.R. Rep. No. 97-420, at 2 (1982).

[33] The Bankruptcy Abuse Prevention and Consumer Protection Act of 2005 (Pub. L. 109-8, 119 Stat. 23) and the Financial Netting Improvements Act of 2006 (Pub. L. 109-390, 120 Stat. 2692).

[34] The following language was added to the definition of commodities, forward, repo, and securities contracts: "any other agreement or transactions referred to" in the definition and "any combination of the agreements or transactions referred to" in the definition.

[35] For example, in the simplest case of two contracts, the non-defaulting firm is owed $1,000 by the bankrupt firm on, say, an interest rate swap (derivative) contract, and owes the defaulting

of differing types to be netted against one another, for example a debt owed on a swap to be netted against a debt owed on an option contract. Netting, whether the netting of like product contracts or cross-product contracts, can reduce the credit exposure of firms that use financial contracts. In turn, the chance that the bankruptcy of one firm might lead to large losses for its financial contract counterparties is reduced, which some observers argue could reduce systemic risk (Jones 1999).[36]

Observers explain that the expansion of special treatment for QFCs occurred in order to account for the considerable growth in the number and diversity of complex financial products over the previous decade (Jones 1999, Skadden 2010). These instruments grew in popularity as they served as mechanisms for financial firms to insure and hedge against risk, helping to reduce uncertainty and stabilize earnings. This increasingly expansive protection for derivatives and repos was intended to achieve the goal of "minimizing the systemic risks potentially arising from certain interrelated financial activities and markets."[37,38]

Some Possible Weaknesses of Bankruptcy's QFC Exemption

The onset of the financial crisis led many observers to reexamine whether this systemic risk rationale was consistent with the events that occurred when financial markets became severely stressed during the recent financial crisis. Therefore, the idea that QFCs should be exempt from the stay was revisited in the lead up to Dodd-Frank and ultimately addressed in the OLA. The systemic risk argument is the prominent justification given by those supporting the expansion of the special treatment given to QFCs. However, there is another cohort, which argues that any reduction in systemic risk, because of QFC exemptions, may be offset by another form of systemic risk

---

firm $800 on a different interest rate swap contract. Under bankruptcy law, the creditor firm may net the two contract debts such that the $800 it owes the defaulting firm is cancelled (netted against the $1,000) and the defaulting firm ends up owing only $200 to the non-defaulting firm. The non-defaulting firm will have to wait for the bankruptcy process to proceed before being repaid any portion of the remaining $200 it is owed. This outcome is superior for the non-defaulting party compared to the case in which netting were not allowed. Here the non-defaulting party would be required to pay the defaulting party the $800 it owed, but wait for the bankruptcy process to be completed before getting any of the $1,000 defaulting party owes it. Of course, in reality, the defaulting firm and the non-defaulting firm are likely to have many contracts outstanding with one another at the time of default, all of which might be netted (Mengle 2010).

[36] This may have magnified the concentration of the derivatives industry according to Bliss and Kaufman (2006, 67–8), who argue that "by explicitly protecting these netting agreements, the 2005 bankruptcy changes reinforced the competitive advantage of the biggest counterparties."

[37] See Jones 1999.

[38] "Immediate termination of outstanding contracts and liquidation of collateral facilitates the acquisition of replacement contracts, reduces uncertainty and uncontrollable risk, improves liquidity and reduces the risk of rapid devaluation of collateral in volatile markets" (Yim and Perlstein 2001, 3).

involving runs on repos[39] and fire sales[40] of the collateral underlying closed-out derivatives contracts (Edwards and Morrison 2005, Taylor 2010, Acharya et al. 2011). The simultaneous termination and liquidation of numerous QFCs (which is allowed by the exemption of QFCs from the stay) may lead to fire sales and possibly further insolvencies. In Lehman's case, of their 930,000 derivatives counterparties, 733,000 sought to terminate their contracts upon their bankruptcy filing on September 15, 2008 (Miller 2009).

Additionally, some observers note that the 2005 bankruptcy laws, which, among other things, extended QFC protections to repos backed by all types of collateral, including all mortgage-related securities, may have encouraged use of mortgage-backed securities as repo collateral (Lubben 2010), and thereby contributed to losses during the financial crisis (Skeel 2010, Government Accountability Office 2011). As Skeel (2010) points out, mortgage values could have spiraled down even more had AIG's counterparties been forced to sell a significant amount of the mortgage-related securities they had posted as collateral on their QFCs (which was avoided when AIG was bailed out).

The idea that QFC fire sales might result from their exemption is not new. In fact, it appears to be what led the Federal Reserve to step in and encourage private firms to come to the aid of Long-Term Capital Management L.P. (LTCM), preventing it from entering bankruptcy (Edwards and Morrison 2005).[41]

As discussed, the bankruptcy process can be long, but among other things, this is intended to give the troubled financial firm and its creditors the time to develop plans to salvage the value of the firm. However, with the exemption from the stay, a large financial firm facing possible default (because of a number of factors, such as a recent credit downgrading or an overall crisis of confidence) has a strong incentive not to file for bankruptcy since doing so would likely trigger simultaneous termination of all QFCs (Skeel and Jackson 2012). Thus, a troubled firm may put it off until the last moment and be forced into a rapid liquidation that significantly depresses values to the detriment of other market participants. These arguments suggest that bankruptcy's current treatment of QFCs may not be optimal.

Observers also find that the special treatment given to QFCs—in order to prevent the perceived systemic risks that arise when these instruments are

---

[39] By "runs on repos" we mean when counterparties, en masse, seize the collateral underlying these deposit-like instruments.

[40] The phrase "fire sale" typically refers to the possibility that the sale of an asset might yield a lower-than-typical price if holders of one type of asset attempt to sell en masse. In comparison, the "typical" (non-fire sale) price will result if sales are distributed over time.

[41] Krimminger (1999, 1) notes that, "[i]n the case of LTCM, the absence of any mechanism under the Bankruptcy Code to 'slow' the liquidation of assets and collateral, [a power granted to the FDIC under the Federal Deposit Insurance Act] and the resulting 'dump' upon the markets, was a key motivation for the pre-insolvency facilitation provided by the Federal Reserve Bank of New York."

subjected to the automatic stay—not only create a different form of systemic risk, but weaken market discipline (Edwards and Morrison 2005, Scott 2011). The special treatment awarded to QFC counterparties in bankruptcy essentially places them ahead of all other creditors in the bankruptcy repayment line, allowing QFC counterparties to get out of their contracts when all other creditors cannot. As a result, their incentive to monitor the debtor prior to bankruptcy and base their pricing and investment decisions on the perceived risk of the counterparty may be significantly reduced, increasing moral hazard (Edwards and Morrison 2005, Roe 2011). It is argued that this leads to market distortions whereby debtors favor short-term repo financing over traditional sources of funding, encouraging a more fragile liability structure (Edwards and Morrison 2005, Skeel and Jackson 2012). For example, at the time of Bear Stearns' failure, a quarter of its assets (approximately $100 billion) were funded by repos (Roe 2011). Roe (2011) suggests that, without the priority given to these instruments in bankruptcy, it is plausible that Bear would have financed a much larger proportion of its assets with longer-term debt, which would have allowed for a more stable funding structure during the financial turmoil.

Some observers who support these arguments maintain that QFCs should be subject to the automatic stay provisions in the Bankruptcy Code, although there are a range of views concerning the length of the stay and whether all QFCs should be treated equally. According to Harvey Miller (2009), lead bankruptcy attorney for the Lehman bankruptcy, the automatic stay, as originally contemplated, is intended to provide a firm with the "breathing space" to find a third party source of liquidity or to carry out an "orderly, supervised wind down of its business assets." Miller argues that, had the special treatment given to QFCs not applied, Lehman's failure may have been avoided and certainly would not have been as "systemically challenging." For instance, Lehman suffered a significant loss of value when nearly 80 percent of their derivatives counterparties terminated their contracts upon their filing of bankruptcy (Miller 2009).

The OLA's One-Day Automatic Stay for QFCs

Given the controversy—with some experts arguing the exemption from the stay is necessary to prevent systemic risk and others arguing that the exemption creates systemic risk—it is natural that Congress chose a solution that leaves the FDIC with discretion to determine the treatment of QFCs for covered financial companies. Under Congress's solution, QFCs are subject to a

one-day automatic stay upon appointment of the FDIC as receiver, whereas QFCs are subject to no stay in bankruptcy.[42]

During the one-day stay under the OLA, the FDIC, as receiver of the failing financial company, must quickly identify how to manage the SIFI's QFC portfolio. The one-day stay is aimed at addressing fears associated with a failing firm's QFC counterparties cancelling their contracts all at once and driving asset prices down. Instead, counterparties' rights to cancel their contracts are put on hold for one day while the FDIC determines how to treat these contracts. The FDIC has this same type of authority when dealing with bank failures. Under the OLA, during this short period, the FDIC has the option to retain the QFCs in receivership, transfer QFCs to another financial institution (to an outside acquirer or to a bridge company created by the FDIC), or reject the QFCs.[43] However, in all instances, the FDIC must retain, reject,[44] or transfer *all* of the QFCs with a particular counterparty and its affiliates.[45,46]

Each action taken by the FDIC has different implications for QFC counterparties of the debtor, as well as the failing firm. Retaining the QFCs in receivership is most similar to bankruptcy in that after the one-day stay expires, QFC counterparties may terminate or net-out their contracts.[47] What differs significantly from bankruptcy, but is very similar to the FDIC's resolution process for depository institutions, is the FDIC's ability to transfer or reject QFCs. If the FDIC chooses to transfer all of the QFCs with a particular counterparty and its affiliates to a third party (including a bridge company), the counterparty is not permitted to exercise its rights to terminate or close out the contract.[48] This awards the FDIC an opportunity to possibly preserve the value of the contracts by removing the ability of counterparties to terminate contracts early and sell off the collateral at fire sale prices (Cohen 2011).

---

[42] The one-day stay lasts until 5:00 p.m. on the business day following the date the FDIC is appointed as receiver. Therefore, the "one-day" stay could last four days if the FDIC is appointed as receiver on a Friday.

[43] For the most part, the FDIC's powers under the OLA to reject or transfer a QFC during their limited one-day stay are much like the powers of the FDIC and bankruptcy trustees under the Federal Deposit Insurance Act and the Bankruptcy Code, respectively, with the exception that they are not supervised by a court nor do they receive counterparty input (Skadden 2010).

[44] In bankruptcy, only contracts or leases that are executory—a contract where both parties have unperformed obligations—may be rejected.

[45] Dodd-Frank Act § 210(c)(9)(A). This is intended to eliminate "cherry picking" (selective assumption and rejection) of QFCs by the debtor.

[46] This differs from the Bankruptcy Code's setoff provision, which allows a creditor to offset all obligations under a single master agreement but not all of the contracts with a single counterparty and its affiliates (Skeel 2010, Cohen 2011). When Lehman filed for bankruptcy, they were a counterparty to 930,000 derivatives transactions documented under 6,120 master agreements (Summe 2011).

[47] If a nondefaulting counterparty has an unsecured claim after terminating a QFC and liquidating any collateral, the claim would then be subject to the same claims process as other unsecured creditors.

[48] If the counterparty were to default at a later time on a separate occasion, they may exercise their close-out rights.

Moreover, a QFC counterparty may find that their contracts are held with a new, and presumably more stable, counterparty or a temporary bridge bank following the one-day stay and, therefore, may have no incentive to terminate (in addition to the fact that it has no ability to terminate), leaving the market undisrupted by their original counterparty's failure while also maintaining what are possibly valuable hedge transactions. Finally, the FDIC may reject (or repudiate) the QFCs of a given counterparty to the debtor, effectively closing them out at the current market value, if they determine that they are somehow burdensome or doing so would otherwise promote orderly administration.[49] However, counterparties may recover, from the FDIC, any damages suffered as a result of the FDIC's rejection of QFCs.[50]

Possible Weaknesses of OLA's One-Day Stay

Some commentators find that the one-business-day stay does not provide the FDIC with sufficient time to identify the potential recipients of the failed firm's derivatives portfolio (Skeel 2010, Bliss and Kaufman 2011, Summe 2011). Given this time constraint coupled with the "all or nothing" approach to the treatment of QFCs (where the FDIC must retain, reject, or transfer all QFCs with a particular counterparty) and the potential systemic risks from its failure to protect a SIFI's QFCs, some suggest that the FDIC is highly likely to transfer all QFC contracts of a given counterparty to a bridge financial institution (i.e., protecting or guaranteeing them in full) (Skeel 2010). After all, if the FDIC does not protect all contracts, then the non-defaulting counterparties may close out and liquidate their contracts upon the expiration of the one-day stay, effectively resulting in the systemic problems previously discussed related to the QFC exemption—closing out the contracts and selling collateral at fire sale prices. Thus, even if various QFC counterparties have differing risk exposures to the defaulting firm, they are all likely to be treated the same and "bailed out." If counterparties believe that their QFCs are likely to be protected by placement in a well-funded bridge company, they are likely to provide more funding (or provide lower-cost funding) to a risky firm than they otherwise would. Further, counterparties may care little about the differing risks associated with the various types of QFCs, because all QFCs of a given counterparty are treated the same. Therefore, while bridge company placement of QFCs may limit systemic risk, it is likely to do so at the cost of increasing moral hazard.

In response to the concern that a one-day stay is likely to lead to the protection of most QFCs, some observers, such as Thomas Jackson, author of a proposal to create a new chapter in the Bankruptcy Code tailored to the

[49] Dodd-Frank Act § 210(c)

[50] Damages are calculated as of the date of repudiation. The word "damages" is defined as the "normal and reasonable costs of cover or other reasonable measures of damages utilized in the industries for such contract and agreement claims" Dodd-Frank Act § 210(c)(3)(C).

resolution of SIFIs (Chapter 14), proposes an extension of the duration of the automatic stay for QFCs to three days. Jackson and others argue that a longer stay duration will give the FDIC additional time to make an informed decision regarding how to handle the failing firm's QFC portfolio (Jackson 2011). Jackson's three-day stay appears to be an attempt to balance the desire to give the FDIC more time, against the danger of producing QFC counterparty failures.[51]

Moreover, the protections for derivatives contracts have broadened over the last several decades and this legislation does not account for the differences across QFC products (such as between repos and swaps), or the types of collateral backing QFCs, which some observers believe should be considered. For instance, several observers find that special treatment (i.e., exemption from the stay) should be limited to derivatives collateralized by highly liquid collateral, such as short-term Treasury securities, since there is little reason to assume that such instruments are important for the going-concern value of the bankrupt firm (Herring 2011, Jackson 2011). In Jackson's 2011 Chapter 14 proposal, highly liquid, or otherwise highly marketable, instruments with no firm-specific value remain exempt from the stay so that creditors who rely on the immediate availability of their funds can get them back quickly and without disruption upon the failure of a firm. On the other hand, the exemption is removed (i.e., the stay would apply) for less liquid instruments, such as CDS, in an effort to prevent these creditors from running on the troubled firm. Clearly, there remains a good bit of controversy about the best way to handle the QFC exemption, in both bankruptcy and the OLA, with no obvious best solution.

3. CONCLUSION

While bankruptcy probably provides the ideal failure resolution mechanism for most corporations, it may not be optimal for some financial firms (i.e., SIFIs). Financial firms are typically more heavily dependent on short-term funding, often including a heavy reliance on QFCs, and their balance sheets are opaque. Because of this dependence on short-term funding, a long stay, while the bankruptcy process plays out, is likely to result in financial difficulties for some of the troubled firm's counterparties. Moreover, DIP funding, which is the usual means of keeping a troubled, but viable, firm alive during reorganization, is likely to be quite difficult to arrange, given the opacity of most financial firms. Because of these weaknesses, handling a SIFI through bankruptcy is likely

[51] While the three-day stay may not provide significantly more time than one day to make such valuations, the Dodd-Frank requirement that SIFIs create resolution plans or "living wills" and provisions forcing swaps to be traded on exchanges could expedite the QFC valuation process, improving the ability of the FDIC to make appropriate decisions within a three-day stay period.

to result in significant risks to financial stability. Policymakers are therefore understandably reluctant to allow SIFIs to enter bankruptcy, given that these risks can be mitigated through bailouts. But bailouts, or the expectation that they could be forthcoming, drive down economic efficiency by exacerbating moral hazard problems.

In an effort to address these difficulties, the OLA was created with the explicit goals of mitigating risk to the financial system and minimizing moral hazard. Specifically, the OLA adjusts the way that QFCs are handled and how creditors are paid out. Despite the attempt to achieve these well-founded goals, because they are conflicting, reducing one inevitably leads to an increase in the other. The one-day QFC exemption does not clearly resolve potential risks to financial stability and it also does not go far to ameliorate the moral hazard problem that is apparent when giving QFCs special treatment. Additionally, the ability to pay some creditors more than they would be likely to receive in bankruptcy may reduce systemic risk, but at the cost of increasing moral hazard. In conclusion, the threat of a SIFI's failure, or the failure itself, presents policymakers with a daunting challenge that neither bankruptcy nor the OLA seems capable of fully resolving.

REFERENCES

Acharya, Viral V., Barry Adler, Matthew Richardson, and Nouriel Roubini. 2011. "Resolution Authority." In *Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance*, edited by Viral V. Acharya, Thomas F. Cooley, Matthew Richardson, and Ingo Walter. Hoboken, N. J.: John Wiley & Sons, Inc., 213–40.

Acharya, Viral, Thomas Cooley, Matthew Richardson, and Ingo Walter. 2009. "Chapter 6, Taxing Too-Big-to-Fail Institutions: The Current Proposals." In *Real-Time Recommendations for Financial Reform*. New York University: The NYU Stern Working Group on Financial Reform (December).

Bernanke, Ben S. 2009. Testimony of Ben S. Bernanke, Chairman of the Board of Governors of the Federal Reserve System, to the U.S. House of Representatives Committee on Financial Services, October 1. Available at www.federalreserve.gov/newsevents/testimony/bernanke20091001a.htm.

Bernanke, Ben S. 2010. Remarks on "The Squam Lake Report: Fixing the Financial System" at the Squam Lake Conference, New York, June 16. Available at www.federalreserve.gov/newsevents/speech/bernanke20100616a.htm.

Bliss, Robert R., and George G. Kaufman. 2011. "Resolving Large Complex Financial Institutions: The Case for Reorganization." Available at www.richmondfed.org/conferences_and_events/research/2011/pdf/bankruptcy_workshop_2011_bliss_paper.pdf.

Bliss, Robert R., and George G. Kaufman. 2006. "Derivatives and Systemic Risk: Netting, Collateral, and Closeout." *Journal of Financial Stability* 2 (April): 55–70.

Boul, Harry. 2006. "Repeat Filings under BAPCPA: Stays, Multiple Discharges and Chapter 20." *ABI Committee News* Volume 4, Number 6 (December). Available at www.abiworld.org/committees/newsletters/consumer/vol4num6/RepeatFilings.html.

Bris, Arturo, Ivo Welch, and Ning Zhu. 2006. "The Costs of Bankruptcy: Chapter 7 liquidation versus Chapter 11 Reorganization." *The Journal of Finance* 61 (June): 1,253–303.

Cohen, Hollace T. 2011. "Orderly Liquidation Authority: A New Insolvency Regime to Address Systemic Risk." University of Richmond *Law Review* 45 (May): 1,143–229.

Duffie, Darrell. 2011. *How Big Banks Fail, and What to Do About It*. Princeton, N. J.: Princeton University Press.

Edwards, Franklin R., and Edward R. Morrison. 2005. "Derivatives and the Bankruptcy Code: Why the Special Treatment?" *Yale Journal on Regulation* 22 (January): 101–33.

Government Accountability Office. 2011. "Bankruptcy: Complex Financial Institutions and International Coordination Pose Challenges." U.S. Government Accountability Office, Report to Congressional Committees, GAO-11-707 (July). Available at www.gao.gov/assets/330/321213.pdf.

Gruenberg, Martin J. 2012. Remarks to the Federal Reserve Bank of Chicago Bank Structure Conference, Chicago, May 10. Available at www.fdic.gov/news/news/speeches/chairman/spmay1012.html.

Herring, Richard J. 2011. The Meeting of the Systemic Resolution Advisory Committee of the Federal Deposit Insurance Corporation, Washington, D.C., June 21. Available at www.fdic.gov/about/srac/11JuneMeetingMins.PDF.

Jackson, Thomas. 2011. "Bankruptcy Code Chapter 14: A Proposal." Hoover Institution Resolution Task Force, February 2012. Available at http://media.hoover.org/sites/default/files/documents/Bankruptcy-Code-Chapter-14-Proposal-20120228.pdf.

Jessup, Clifton R., Jr. 1995. "Should the Automatic Stay Be Abolished?" The ABI National Symposia Series, ABI Bankruptcy Reform Study Project. Alexandria, Va.: American Bankruptcy Institute (June). Available at www.abiworld.org/abidata/online/symposium/stay.html.

Jones, Douglas H. 1999. "Statement of Douglas H. Jones, Senior Deputy General Counsel, Federal Deposit Insurance Corporation, on Bankruptcy Reform Legislation, Committee on Banking, United States Senate, March 25, 1999." Available at www.fdic.gov/news/news/speeches/archives/1999/sp25mar99b.html.

Krimminger, Michael. 1999. "Insolvency in the Financial Markets: Banks, Hedge Funds and Other Complications." *Banking Policy Report*, January 18.

Lubben, Stephen J. 2010. "The Bankruptcy Code Without Safe Harbors." *American Bankruptcy Law Journal* 84: 123–38.

Mengle, David. 2010. "The Importance of Close-Out Netting." *ISDA Research Notes* Number 1.

Miller, Harvey R. 2009. "Too Big to Fail: The Role for Bankruptcy and Antitrust Law in Financial Regulation Reform." Testimony before the Subcommittee on Commercial and Administrative Law, House of Representatives, Committee on the Judiciary, October 22.

Morgan, Donald P. 2002. "Rating Banks: Risk and Uncertainty in an Opaque Industry." *American Economic Review* 92 (September): 874–88.

Roe, Mark J. 2011. "The Derivatives Market's Payment Priorities as Financial Crisis Accelerator." *Stanford Law Review* 63 (March): 539–90.

Scott, Kenneth. 2011. "A Guide to the Resolution of Failed Financial Institutions: Dodd-Frank Title II and Proposed Chapter 14." Stanford Law School. Available at http://media.hoover.org/sites/default/files/documents/ken-scott-guide-to-resolution-project.pdf (May).

Skadden. 2010. "The Dodd-Frank Act: Commentary and Insights." Available at www.skadden.com/sites/default/files/publications/Skadden_Insights_Special_Edition_Dodd-Frank_Act1_6.pdf.

Skeel, David A. 2010. "Bailouts, Bankruptcy or Better? Resolution Authority II." In *The New Financial Deal: Understanding the Dodd-Frank Act and its (Unintended) Consequences*. Hoboken, N. J.: John Wiley & Sons, Inc., 129–52.

Skeel, David A., and Thomas Jackson. 2012. "Transaction Consistency and the New Finance in Bankruptcy." *Columbia Law Review* 112 (February).

Summe, Kimberly. 2011. "An Examination of Lehman Brothers' Derivatives Portfolio Post-Bankruptcy and Whether Dodd-Frank Would Have Made Any Difference." Available at http://media.hoover.org/sites/default/files/documents/Kimberly-Summe-Dodd-Frank-20110421.pdf.

Taylor, John B. 2010. "Defining Systemic Risk Operationally." In *Ending Government Bailouts as We Know Them*, edited by George Shultz, Kenneth Scott, and John B. Taylor. Stanford University: Hoover Institution Press, 33–57.

White, Michelle J. 1998. "Corporate Bankruptcy." In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman. Macmillan Press. Available at http://weber.ucsd.edu/~miwhite/.

Yim, Soo J., and William J. Perlstein. 2001. "The Effect of Proposed Amendments to U.S. Insolvency and Banking Laws on Transactions Involving Securities, Commodities and Other Financial Contracts." Presentation prepared in connection with the American Bar Association 2001 Spring Meeting Section of Business Law Business Bankruptcy Committee Forum on Derivatives And Proposed Financial Contract Netting Legislation, March 24.

Economic Quarterly—Volume 98, Number 1—First Quarter 2012—Pages 33–50

Contingent Capital: The Trigger Problem

Edward Simpson Prescott

Contingent capital is long-term debt that automatically converts to equity when a trigger is breached. It is a new and innovative security that many people are proposing as part of a reform in bank capital regulations.[1] The security is most associated with Flannery (2005), but with the recent financial crisis many others, including Flannery (2009); Huertas (2009); Albul, Jaffee, and Tchistyi (2010); Plosser (2010); Squam Lake Group (2010); Calomiris and Herring (2011); McDonald (2011); Pennacchi (2011); and Pennacchi, Vermaelen, and Wolff (2011), have also advocated its adoption.[2] Furthermore, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 mandated a study of contingent capital, while the Independent Commission on Banking's report (2011) on banking in the United Kingdom recommended that bank capital structure include loss-absorbing debt like contingent capital.

Contingent capital has four appealing properties. First, it increases a bank's capital when a bank is weak, which is precisely when it is hardest for a bank to issue new equity. In doing so, contingent capital reduces the "debt overhang" problem, which is the inability of a bank to raise funds to finance new loans because their return partially accrues to existing debtholders. During the recent financial crisis, many U.S. banks were forced to raise new equity. If they had had contingent capital securities, this process would have been much easier. Second, contingent capital automatically restructures part

The views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: edward.prescott@rich.frb.org.

[1] There are already several changes to bank capital requirements, like increasing capital levels and making them more procyclical, that have already been implemented or are in the process of being implemented.

[2] See Calomiris and Herring (2011) for a more detailed list of the various contingent capital proposals. An early analysis is contained in Raviv (2004). Finally, there are also related proposals like Hart and Zingales (2011) that require banks to raise more equity when price triggers are breached.

of a bank's capital structure, reducing the chance it fails and is put in resolution or bankruptcy.[3] Many people think that the abrupt nature of Lehman's bankruptcy was very disruptive to financial markets, so a pre-bankruptcy reorganization of a financial firm may be valuable. Third, it is a way to force regulators to act, at least when the trigger is tied to an observable variable, like the price of a bank's equity. Fourth, it "punishes" equityholders by diluting existing equityholders. Some proposals argue that the threat of this dilution will give a bank an incentive to take less risk (e.g., Calomiris and Herring 2011).

All four of these properties have varying degrees of merit, but the purpose of this article is not to analyze these benefits.[4] Instead, it is to discuss a cost of implementing contingent capital. All contingent capital proposals rely on a trigger to implement conversion. Many of the proposals advocate the use of a market-price trigger (e.g., Flannery 2005, 2009), but some of them rely on accounting numbers (e.g., Huertas 2009), and others also include a role for regulators. For example, Squam Lake Group (2010) advocates as a trigger the use of accounting numbers at the firm level plus a regulatory declaration that there is a systemic crisis.

This article argues that the trigger is the weak point of contingent capital and, more specifically, a trigger based on a market price, be it a fixed trigger or a signal for a regulator to act, suffers from an inability to price contingent capital. This inability will be more precisely defined later, but the problem arises because asset prices incorporate the possibility of conversion and the way in which contingent capital conversion is triggered makes this feedback problematic. In practice, this will mean conversion could occur when it is not desired. Unless the price trigger can be designed in a way to overcome this problem, contingent capital with a price trigger will not work.

An alternative to a price-based trigger is an accounting-based one. This article does not focus on this type of trigger except to note that accounting measures of a bank's quality seem to lag its actual condition. For example, the prompt corrective action provision (PCA) of the Federal Deposit Insurance Corporation (FDIC) Improvement Act of 1991 is an accounting-based regulatory trigger system. It does not convert debt to equity like with contingent capital, but it requires regulators to restrict the activities of a bank and even shut the bank down if regulatory capital drops below certain thresholds. The motivation behind PCA was to force regulators to act before a bank's losses got too big. In the recent crisis, losses to the deposit insurance fund have been very high despite the existence of PCA (Government Accountability Office

[3] It is worth noting that even though converting debt to equity raises the book value of equity, it does not bring new cash into a firm (other than indirectly by eliminating interest payments on the converted debt) like a new issuance of equity would.

[4] This is not entirely true. The Appendix contains a discussion of why the incentive effects of contingent capital are not the major benefit of contingent capital.

2011). For example, FDIC losses on banks and thrifts (excluding Washington Mutual) that failed over the period 2007–2010 have been 24.62 percent of the assets of these failed institutions.[5] Based on this experience, caution about the timeliness of accounting measures seems warranted.

Underlying the use of price triggers in contingent capital is the fundamental idea that prices aggregate information, so regulation should be able to use them to make decisions. Indeed, one of the most robust findings in financial economics is that prices are efficient in the sense that prices incorporate all available information (Fama 1970). A striking example is found in Roll (1984), who documents that the price of orange juice futures better predicts variations in Florida weather than National Weather Service forecasts. Indeed, the empirical banking literature surveyed in Flannery (1998) documents that bank security prices can predict changes in supervisory ratings.[6]

This article uses a simple model to illustrate how the usual theoretical and empirical properties of financial prices break down for contingent capital with a price trigger. The model is based on a small theoretical literature that has found that the discrete jump in security prices resulting from conversion interferes with the ability of prices to aggregate information.[7] This problem with contingent capital was discovered by Sundaresan and Wang (2011), who found that contingent capital with a trigger based on an equity price could not be priced because there did not necessarily exist a unique set of prices. When conversion heavily diluted equity, they found that there were multiple equilibria. When conversion raised the value of equity, they found that there were no equilibria.

Birchler and Facchinetti (2007) and Bond, Goldstein, and Prescott (2010) studied the related problem of a regulator who could intervene in the operations of a bank and thus affect the value of the bank. In both articles, the regulator did not know the fundamental value of the bank, but instead had to infer it from the prices of the traded bank securities. Instead of using a price-trigger rule, the regulator had trigger-like preferences in that he wanted to intervene only when the fundamental quality of the bank was below some threshold. The

---

[5] Washington Mutual is excluded for two reasons. First, including it skews the average because it had about $300 billion in assets and the FDIC took no loss on it when they arranged a sale through receivership to J.P. Morgan Chase. The high average on the rest of the failed banks illustrates that there were a lot of banks for which the accounting numbers substantially lagged their actual condition, otherwise losses would have been much smaller. Second, Washington Mutual was not shut down because of a violation of PCA triggers, but rather because of liquidity problems. Indeed, it was well capitalized by PCA standards as of September 25, 2008 (Offices of Inspector General 2010), so it is further evidence that accounting numbers can lag actual condition.

[6] Motivated by this logic, there is an older set of proposals (e.g., Stern 2001) that advocate that bank supervisors use market prices to supplement their surveillance of banks.

[7] One concern raised about the use of market price triggers is that traders will manipulate prices to generate conversion when it would be advantagous to them. While this is a legitimate concern, the analysis in this article shows that there are problems with using market prices as a trigger even in the absence of these concerns.

effect of the intervention decision is mathematically similar to the effect from a price trigger—there is nonexistence of equilibrium when the regulator cannot commit to an intervention rule, though in the simplest environments there is a unique equilibrium when there is heavy dilution. Indeed, the implication of their work is that when prices are used as a trigger, prices need *not* aggregate all available information.

Almost all the analysis of contingent capital is theoretical because there is no financial market evidence. Sundaresan and Wang (2011) report only four issuances of contingent capital, all of which were within the last few years. Furthermore, none of these issuances purely rely on market prices. For example, Credit Suisse (2011) issued a contingent-capital security in 2011 that used as its trigger the equity capital ratio and allowed the regulator to trigger conversion if it was determined that customary measures to improve capital adequacy were inadequate to keep Credit Suisse viable.

To overcome this lack of data, Davis, Korenok, and Prescott (2011) ran market experiments to study the effect of using a market price as a contingent capital trigger. Market experiments are small scale economies run in laboratories with human subjects who trade in a market. They found that conversion increased the volatility of prices, reduced the efficiency of allocations, and led to conversion errors with some frequency. A summary of their findings is provided.

Section 1 illustrates the pricing problem with a simple theoretical model. Section 2 discusses possible ways around the pricing problem. Section 3 briefly discusses the experimental results. Section 4 concludes, and the Appendix contains an argument for why contingent capital will only partially reduce risk-taking incentives.

1. THE MODEL

There is a bank that is financed by one unit of equity and one unit of debt. Debt is scheduled to pay one and there is one share of equity. The value of the bank, that is, the amount of cash it has to distribute, is $\theta > 0$.

The bank's equity is traded in a market by risk-neutral traders. These traders know the value of θ and use that information plus their expectation of whether debt will be converted to equity to trade the equity. The price of equity depends on θ and is written $p(\theta)$.

For simplicity, this article only considers conversion rules in which all the debt is converted to equity. This assumption is not important for the results. The *conversion rule* is $\alpha(p)$, which at price p converts the single unit of debt into α shares of equity. As with the trigger rule proposals, the conversion depends on the price of equity. There are a lot of possible conversion rules, but the most common ones are to convert the debt to a fixed number of shares.

In particular, they take the form

$$\alpha(p) = \begin{cases} \alpha > 0 & \text{if } p \leq \hat{p} \\ 0 & \text{if } p > \hat{p} \end{cases},$$

where $\hat{p}$ is some fixed trigger. The idea is that as a bank gets closer to insolvency, its share price will drop and that is when it is best to automatically convert debt to equity.

Definition 1 *Given a trigger rule, $\alpha(p)$, an equilibrium is a price of equity, $p(\theta)$, such that $\forall\theta$*

$$p(\theta) = \begin{cases} \frac{\theta}{1+\alpha(p(\theta))} & \text{if } \alpha(p(\theta)) > 0 \\ \theta - 1 & \text{if } \alpha(p(\theta)) = 0 \end{cases}. \tag{1}$$

Equilibrium requires that prices, $p(\theta)$, be consistent with the conversion rule. As we will see, for some conversion rules no $p(\theta)$ will satisfy (1) and for others multiple $p(\theta)$ will.

No Conversion Benchmark

As a benchmark, consider the case of no conversion of debt. In this case, the price of equity is

$$p(\theta) = \begin{cases} 0 & \text{if } \theta \leq 1 \\ \theta - 1 & \text{if } \theta > 1 \end{cases}.$$

When $\theta \leq 1$, all the firm's payments go to the debtholders and there is nothing left for equityholders. When $\theta > 1$, the debtholders get the full payment of one and the equityholders get what is left.

Decreased Value of Equity

Most contingent capital proposals advocate setting conversion so as to heavily dilute equity in order to "punish" the owners of the bank.[8] The problem with a trigger rule that heavily dilutes equity is that there are multiple equilibria. To illustrate the problem, consider the trigger rule that if the price of equity is less than or equal to 1.5 then the debt is converted to one share of equity, so there are two shares of equity total. Formally,

$$\alpha(p) = \begin{cases} 1 & \text{if } p \leq 1.5 \\ 0 & \text{if } p > 1.5 \end{cases}. \tag{2}$$

Under this trigger rule, an equilibrium exists. One of them is

$$p(\theta) = \begin{cases} \theta/2 & \text{if } \theta \leq 3 \\ \theta - 1 & \text{if } \theta > 3 \end{cases}.$$

[8] See the Appendix for a discussion of incentives for equity owners.

To see this, if, at $\theta \leq 3$, the traders assume that there will be conversion, then the price is less than or equal to 1.5, which is consistent with the conversion rule. Similarly, for $\theta > 3$, if the traders assume that there is no conversion, then the price is $\theta - 1 > 1.5$, which is also consistent with the conversion rule.

A second equilibrium is

$$p(\theta) = \begin{cases} \theta/2 & \text{if } \theta \leq 2.5 \\ \theta - 1 & \text{if } \theta > 2.5 \end{cases}.$$

At $\theta \leq 2.5$, if traders assume there will be conversion, then the price will be less than or equal to 1.25, which is consistent with the conversion rule. Similarly, for $\theta > 2.5$, if the traders assume that there is no conversion, then the price is $\theta - 1 > 1.5$, which is also consistent with the conversion rule.

As should be apparent, any price function in which traders assume that there will be conversion for values of θ below any cutoff between 2.5 and 3.0 will be an equilibrium. But actually, the multiple equilibria problem is even worse than this. There are *lots* of other price functions that are equilibria, some of which are rather strange. For example,

$$p(\theta) = \begin{cases} \theta/2 & \text{if } \theta \leq 2.5 \\ \theta - 1 & \text{if } 2.5 < \theta \leq 2.6 \\ \theta/2 & \text{if } 2.6 < \theta \leq 3 \\ \theta - 1 & \text{if } \theta > 3 \end{cases}$$

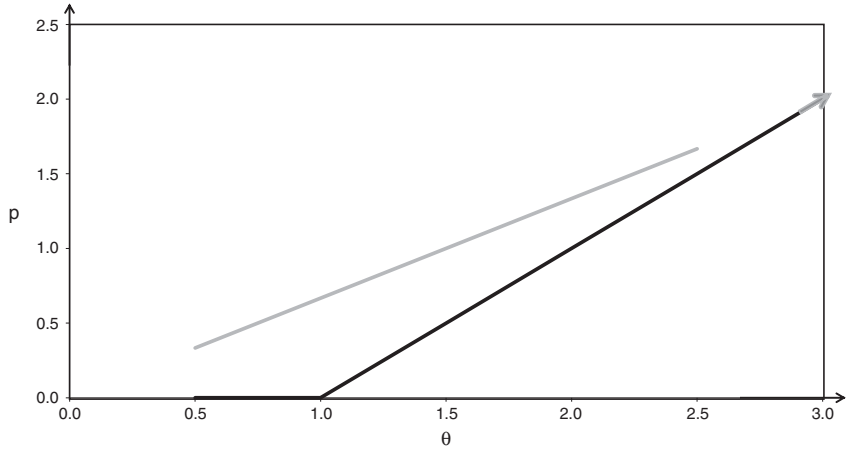
is also an equilibrium!

Multiple equilibria is a serious problem for contingent capital because it is unclear what its price will be. As we will see, a variety of prices occur in the experimental evidence. In terms of the proposal this means that conversion need not happen when it is desired or it may happen when it is undesired.

Increased Value of Equity

The proposals do not advocate conversion to increase the value of equity, but this case still has to be studied for two reasons. First, there may very well be states of the world where the price of equity is low, but conversion would increase the value of equity. For example, imagine a very high probability that θ will be less than 1, the amount owed to debtors. Equity does not have much value in this case, but if the debt is converted to equity, then the price of equity may very well go up even if it is heavily diluted. After all, a high probability of a small payment can be more valuable than a low probability of a high payment. Second, the proposals for regulators to use prices to take regulatory actions, like replacing management or doing something similar, could very well *increase* the value of the bank. This was the scenario studied in Birchler and Facchinetti (2007) and Bond, Goldstein, and Prescott (2010).

Figure 1 Increased Value of Equity Case



Notes: The black line shows the price of equity assuming that the debt is not converted to equity. The gray line shows the price of equity assuming that it converts to equity when $\theta \leq 2.5$. The gray line is non-monotonic, which is suggestive as to why there is no equilibria when the price trigger is set at 1.5. For θ just below 2.5, the price drops below 1.5 without conversion and increases above 1.5 with conversion. Neither possibility is consistent with the trigger rule.

If the value of equity increases from a conversion then the problem is not one of multiple equilibria, but instead that no equilibrium even exists. To see this, consider the same price trigger level as above, but now convert debt into 0.5 shares, that is,

$$\alpha(p) = \begin{cases} 0.5 & \text{if } p \leq 1.5 \\ 0 & \text{if } p > 1.5 \end{cases} . \tag{3}$$

Under this trigger rule, no equilibrium exists. To see this, consider what the price can be if $\theta = 2.5$. If traders assume there will be conversion, then there is no debt and 1.5 shares of equity. The price of equity would then have to be $\frac{2.5}{1.5}$, but that is greater than the 1.5 trigger, so there cannot be conversion. Alternatively, if traders assume that there will not be conversion then the price of equity is 1.5 without conversion, but that violates the trigger rule of converting when the price is less than or equal to 1.5.[9]

Figure 1 illustrates the problem. The gray line shows what prices would be *if* conversion could be tied directly to the fundamental value θ. The problem

[9] This is not just a problem right at the trigger point. The same logic applies to a range of fundamentals below 2.5, in this example, down to 2.25.

here is that a conversion rule that increases the price of equity requires a price function that is above the trigger value for a range of θ values below $\theta = 2.5$. This non-monotonicity in prices around the trigger implies that the trigger rule, as commonly proposed, cannot distinguish between values of θ for which conversion is desirable and values for which it is not.

2. SOLUTIONS?

The lack of existence of a unique equilibrium is a serious challenge to implementing contingent capital proposals. Certainly, triggers of the form analyzed above would not work. There are, however, alternative ways to structure the trigger that avoid these problems. Below, some possible solutions are described and assessed.

Getting the Conversion Ratio Just Right

If conversion is set so that the value of equity does not change at conversion, then there is a unique equilibrium. In the example above, a trigger rule that works is at a price of 1.5, convert the debt to $\frac{2}{3}$ a share. Under this rule, if conversion occurs at $\theta = 2.5$, then the price of equity is 1.5, just like if there is no conversion. Figure 2 illustrates.

More generally, the conversion ratio that generates a unique equilibrium is the one that generates a continuous monotonic price function. With a conversion rule that converts all the debt to α shares of equity (like in the examples above), α needs to be set so that at the desired conversion point, $\hat{\theta}$, the prices of equity under conversion and non-conversion are the same, that is,

$$\hat{\theta} - 1 = \frac{1}{1+\alpha}\hat{\theta}$$

or

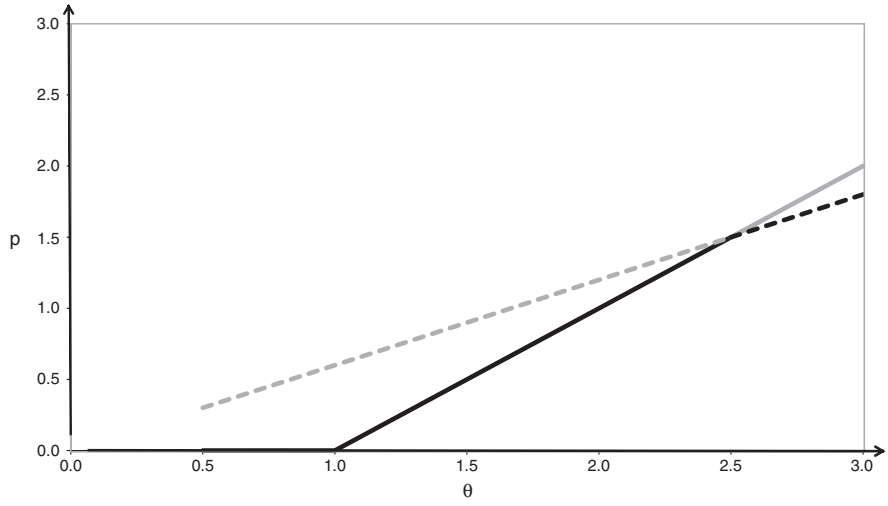
$$\alpha = \frac{1}{\hat{\theta} - 1}.$$

This means that the trigger price in turn needs to be

$$\hat{p} = \frac{1}{\alpha}.$$

While this conversion rule is simple and works in this one-period model, it need not work in a dynamic model with uncertainty. Sundaresan and Wang (2011) show that, in a dynamic model, even if the conversion ratio is set so that at maturity there is no change in the value of equity from conversion that is no guarantee that the same conversion ratio will not change the value of equity in earlier periods. Basically, a simple trigger rule is not robust enough to cover the wide variety of paths of uncertainty.

Figure 2 No Change in the Value of Equity at Conversion



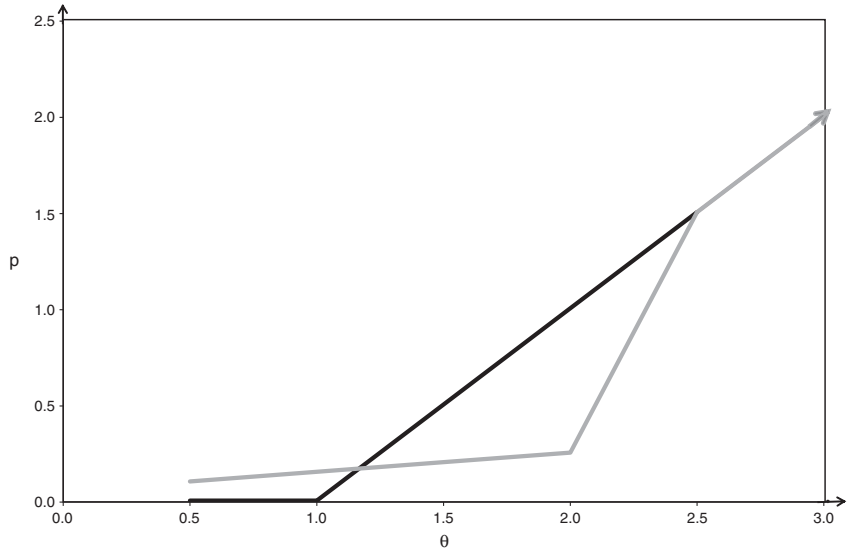
Notes: The solid line shows the price of equity assuming that the debt is not converted to equity. The dashed line shows the price of equity assuming that the debt always converts to equity. The portions of the lines in gray show the price of equity when the conversion ratio is set so that there is no change in the value of equity at the trigger. This price function is the only equilibrium.

Finally, in order to prevent a jump in the value of equity, this conversion rule actually helps the original equity owners, at least relative to no conversion. As Figure 2 illustrates, for values of θ less than 2.5, the price of a share is more than it would be without conversion. With this conversion rule, equity owners are actually not punished, which is one of the motivations behind contingent capital.

Sliding Conversion Rules

One way to "get the conversion ratio just right" without rewarding equity owners is to use a "sliding conversion rule." The idea is to make the amount of dilution vary so that as θ declines, the price continuously decreases. The monotonicity is needed for existence and the continuity is needed for uniqueness. Birchler and Facchinetti (2007) use a similar concept in their regulatory action model to get existence when there is a value-increasing action.

Figure 3 Trigger Rule with Unique Equilibrium that Punishes Equityholders



Notes: The black line shows the price of equity with non-convertible debt. The gray line shows the price of equity with debt that converts to equity using the sliding conversion rule in the text. The price function is continuous and monotonically increasing. Relative to non-convertible debt, it hurts original equity owners for medium values of θ, but helps them for low values of θ.

For this example, assume that the lower bound on θ is 0.5. A conversion function that generates a unique price function is

$$\alpha(p) = \begin{cases} (9p - 0.5)/p & \text{if } 0.10 \leq p \leq 0.25 \\ (4.75 - 1.5p)/2.5p & \text{if } 0.25 < p \leq 1.5 \\ 0 & \text{if } p > 1.5 \end{cases} .$$

Figure 3 shows the price function that results from this conversion rule. It is the piecewise linear gray line, and it is straightforward to show that it is the unique price function. There are three things to note about this function. First, the continuity prevents the multiple equilibria that arose in the heavy dilution example. Second, the monotonicity prevents the discrete drop in price at and above a trigger point, which was the source of nonexistence in the increased value example. Third, the price schedule punishes equity owners for a range of values of θ below the trigger. For the lowest values of θ, equityholders are actually made better off by conversion, but this feature is only there to keep the price of equity above zero. With a trigger and a conversion rule that

wipes out old equity—bankruptcy can be thought of a conversion rule with $\alpha = \infty$—the price of equity is zero under conversion, so there is a really severe form of multiple equilibria, namely, for *any* value of θ there is an equilibrium where the price of equity is zero! Keeping the price of equity positive under conversion prevents this perverse problem.

Use Other Information

Another possible solution is to make the conversion depend on the total value of the firm (e.g., Raviv [2004] and Pennacchi [2011]). In the example, the total value of the firm is simply the value of equity plus the value of contingent capital. If the trigger were set so that the total value of the firm is less than or equal to 2.5, then a unique equilibrium would exist. The reason is that the value of equity plus debt is simply the value of the firm, that is, the cash flow θ, and that does not change with conversion.

The obvious concern with this solution is that markets for bank debt (not to mention bank deposits) are far less liquid than those for equity, so good measures of the value of the firm will not be readily available. But even if this issue could be overcome, the deeper issue is whether conversion affects the value of a firm. The firm-value trigger works in this example because the model is a Modigliani-Miller environment in that the capital structure does not affect the value of the firm. However, implicit in many of the arguments for contingent capital is that a debt-to-equity conversion will *improve* the value of the firm by reducing debt overhang. But if there is a debt overhang problem then the environment is *not* a Modigliani-Miller one, so a change in the capital structure would create a discrete change in the value of the firm and there would be the same problems with equilibrium that we analyzed above.[10]

Price Restrictions

A simple way to deal with the multiple equilibria is to forbid exchanges of equity at certain prices. In the decreased value of equity example above, if equity were forbidden to trade over the range $(1.25, 1.5]$ then the only equilibrium would be the one where conversion occurs for $\theta \leq 2.5$. The other equilibria discussed above simply cannot occur.

Even if it were feasible to prohibit trading at certain prices, this solution would still require a lot of information to set up. The amount of the drop in the price of equity will depend on the aggregate state (something that was not in

[10] The Birchler and Facchinetti (2007) and Bond, Goldstein, and Prescott (2010) studies were precisely worried about regulatory interventions that changed, and more specifically improved, the value of the bank.

the model above). That requires a lot of information on the part of regulators to set up.

Prediction Markets

Another possible solution is to introduce prediction markets in whether or not there is conversion and use that information as part of the trigger. Bond, Goldstein, and Prescott (2010) show that in the regulatory action with an increased value of equity case, when prediction markets are added, a unique equilibrium exists. Here, we show that with a price trigger rule that also depends on the price of the prediction security, a unique equilibrium exists for both the decreased and increased value examples.

The prediction market is a market in a security that pays one if there is conversion and zero otherwise. The same traders who trade equity also trade the prediction security. The price of the prediction security is $q(\theta)$ and the trigger rule now depends on both prices, that is, $\alpha(p, q)$. A price of one means that traders expect conversion and a price of zero means they do not.

Definition 2 *Given a trigger rule, $\alpha(p, q)$, an equilibrium is a price of equity, $p(\theta)$, and a price of the prediction security, $q(\theta)$, such that $\forall\theta$*

$$p(\theta) = \begin{cases} \frac{\theta}{1+\alpha(p(\theta),q(\theta))} & \text{if } \alpha(p(\theta), q(\theta)) > 0 \\ \theta - 1 & \text{if } \alpha(p(\theta), q(\theta)) = 0 \end{cases}$$

$$q(\theta) = \begin{cases} 0 & \text{if } \alpha(p(\theta), q(\theta)) = 0 \\ 1 & \text{if } \alpha(p(\theta), q(\theta)) > 0 \end{cases}.$$

For the example studied earlier, where the value of equity declines with conversion, consider the following modification to the trigger rule (2):

$$\alpha(p, q) = \begin{cases} 1 & \text{if } p \leq 1.25 \\ 1 & \text{if } 1.25 < p \leq 1.5 \text{ and } q = 0 \\ 0 & \text{if } 1.25 < p \leq 1.5 \text{ and } q = 1 \\ 0 & \text{if } p > 1.5 \end{cases}.$$

The price function

$$p(\theta) = \begin{cases} \theta/2 & \text{if } \theta \leq 2.5 \\ \theta - 1 & \text{if } \theta > 2.5 \end{cases}$$

$$q(\theta) = \begin{cases} 1 & \text{if } \theta \leq 2.5 \\ 0 & \text{if } \theta > 2.5 \end{cases}$$

is an equilibrium. For $\theta \leq 2.25$, conversion has to happen, while for $\theta > 3$, conversion cannot happen. Where the prediction security gets used is for the range of θ where multiple equilibria were an issue without the prediction security. First, consider the range $2.25 < \theta \leq 2.5$. If traders assume that there will be no conversion, then $1.25 < p \leq 1.5$ and $q = 0$, but by the trigger

rule there will be conversion. If traders assume there will be conversion, then $p \leq 1.25$, and there is conversion (and $q = 1$), which is consistent with the trigger rule. Second, consider the range $2.5 < \theta \leq 3$. If traders assume that there will be conversion, then $1.25 < p \leq 1.5$ and $q = 1$, but by the trigger rule there will not be conversion. In contrast, if traders assume there will not be conversion, then $p > 1.5$, and there is no conversion (and $q = 0$), which is consistent with the trigger rule. This trigger rule eliminates the multiple equilibria by making it impossible for prices to fall in the range between 1.25 and 1.5, which prevents conversion at values of $\theta > 2.5$. Essentially, this solution uses the trigger rule to restrict the prices in the same way that the analysis of the price-restriction solution did earlier.

In the case where the value of equity increases, where existence of equilibrium was the problem earlier, the prediction market gives the trigger rule enough extra information to recover existence. Consider the modification to the earlier trigger rule (3):

$$\alpha(p,q) = \begin{cases} 0.5 & \text{if } p \leq 1.5 \\ 0.5 & \text{if } 1.5 < p \leq 1\frac{2}{3} \text{ and } q = 1 \\ 0 & \text{if } 1.5 < p \leq 1\frac{2}{3} \text{ and } q = 0 \\ 0 & \text{if } p > 1\frac{2}{3} \end{cases}.$$

The price function

$$\begin{aligned} p(\theta) &= \begin{cases} \frac{2}{3}\theta & \text{if } \theta \leq 2.5 \\ \theta - 1 & \text{if } \theta > 2.5 \end{cases} \\ q(\theta) &= \begin{cases} 1 & \text{if } \theta \leq 2.5 \\ 0 & \text{if } \theta > 2.5 \end{cases} \end{aligned}$$

is a unique equilibrium. To see this, first consider $\theta \leq 2.5$. If traders assume conversion, then $p \leq 1\frac{2}{3}$ and $q = 1$, which is consistent with the trigger rule. If traders assume no conversion then $p \leq 1.5$, but that requires conversion according to the trigger rule, so that is not a possibility. Now consider $\theta > 2.5$. If traders assume that there is no conversion, then $p > 1.5$ and $q = 0$, which is consistent with the trigger rule. However, if traders assume conversion, then $p > 1\frac{2}{3}$, which by the trigger rule requires no conversion, so that is not a possibility.

A prediction market security of the form discussed above does not exist right now. Nevertheless, credit default swaps are very close in that they are essentially insurance contracts that pay out in the event of a default. If a credit default swap was designed so that conversion was the triggering "default" event, then the swap could be used as the prediction security. Of course, the usual concerns about liquidity and market manipulation would apply.

3. EVIDENCE

As was discussed earlier, there is very little empirical evidence on the effectiveness of contingent capital. The only source of evidence that I am aware of is from the laboratory experiments reported in Davis, Korenok, and Prescott (2011). Laboratory experiments are games played by subjects (typically college students) for real stakes. The experiments can be used to study individual decisionmaking or more complex group interactions.

Davis, Korenok, and Prescott (2011) ran experiments where the subjects used a standard open book double auction to trade an asset that could change in value if a price trigger were breached. The price trigger worked just like the examples above. If breached, the underlying value of the asset jumped up in some of the experiments and dropped in others.

As predicted by theory, they found that the fixed-price trigger created informational inefficiencies in the sense that prices deviated from fundamentals and were more volatile. This was true in experiments where the value of equity was increased and those where it was decreased. The problems were worse, however, in the case where the value increased.[11]

Compared with a no-conversion baseline, they also found that conversion made the allocation less efficient in the sense that assets ended up less frequently in possession of the traders who valued them the most. Finally, they also found the trigger was frequently breached when the fundamentals did not warrant conversion and it was sometimes not breached when fundamentals warranted conversion. For some ranges of fundamentals, these errors occurred most of the time. There were some caveats to their findings. In particular, conversion errors in the decreased-value experiments were concentrated in the range of fundamentals just above the trigger, which may be tolerable from a cost-benefit perspective, but for increased-value experiments conversion errors were dispersed over a wider range of fundamentals.

They also ran experiments where, instead, a regulator made the decision of whether or not to convert. The regulator was given a reward structure that rewarded him if he converted when the fundamental was below the trigger or if he did not convert when it was above the trigger. Compared with the fixed-price trigger, performance by a regulator tended to be worse. In particular, it seemed that the additional source of uncertainty for traders, namely, guessing how the regulator would interpret the price, made prices more volatile. Furthermore, the regulator made conversion errors over a wider range of fundamentals than in the fixed-trigger experiments. They also ran experiments with a prediction market in whether the regulator would convert. The additional information from the prediction market improved the efficiency of

[11] This was the case where an equilibrium did not exist in the model.

prices and allocations as well as the performance of the regulator, but substantial inefficiencies remained.[12]

For more details see the article, but overall they concluded that the inefficiencies and frequency of conversion errors are a significant cost to using contingent capital with a price trigger.

4. CONCLUSION

This article illustrates the potential pitfalls of using a market-price trigger in contingent capital. The multiple equilibria and nonexistence results are problematic for these proposals. Indeed, in the closest thing we have to empirical evidence—the market experiment data—the use of a trigger made prices and allocations less efficient, increased volatility, and led to numerous conversion errors.

In my view, any contingent capital proposal that uses market-based prices needs to confront these problems. A viable proposal needs to find a trigger that is not subject to multiple equilibria and nonexistence or, alternatively, one that leads to few conversion errors, minor inefficiencies, and reasonable levels of price volatility.

APPENDIX: A DIGRESSION ON INCENTIVES

Many of the proposals advocating contingent capital emphasize the value of "punishing" the equity owners by diluting equity (e.g., Calomiris and Herring 2011) in order to improve equity owners' ex ante incentives. Structuring bank capital to improve incentives is an idea with a long tradition in the banking literature. The banking literature that came out of the savings and loan crisis emphasized the risk-shifting incentives that bank equity owners have under a legal and regulatory system that includes limited liability and deposit insurance (e.g., White 1991).

This perspective is one that I am sympathetic with, but if incentives are the motivation behind contingent capital, then the analysis is better served by directly using an incentive model with an explicit treatment of moral hazard. The standard approach to analyzing incentives is to use a moral hazard model where bank equity owners have limited liability and can choose the amount

[12] They did not run experiments where a prediction market was combined with a fixed-price trigger.

of risk the bank takes.[13] Interestingly, in this class of models, Marshall and Prescott (2001, 2006) found that the most effective way to discourage a bank from taking excessive risk was to, counterintuitively, "punish" the bank when it did well! (In their context, punishment meant requiring that the bank's capital structure include warrants with a high strike price that essentially reduced the upside gain to the bank. For a summary of their argument, see Prescott [2001].) The reason for their surprising result was that very high returns were more likely when a bank took an excessive amount of risk than an appropriate amount, so reducing equityholders' payoffs in these states was desirable. In their model, it was also desirable to "punish" the equityholders when the bank did poorly, but limited liability reduced the amount of punishment that could be provided in this case.

The point of this digression is to argue that bank incentives need to be viewed from a broad perspective that may well put little emphasis on "punishing" equityholders when a bank does poorly, or more accurately, that the incentive implications of a heavy dilution are only a part, and possibly a small part, of the total incentives created by a bank's capital structure. For this reason, I think recapitalization effects rather than any incentive effects are what is potentially most valuable about contingent capital.

REFERENCES

Albul, Boris, Dwight M. Jaffee, and Alexei Tchistyi. 2010. "Contingent Convertible Bonds and Capital Structure Decisions." Manuscript, University of California at Berkeley.

Birchler, Urs, and Matteo Facchinetti. 2007. "Self-Destroying Prophecies? The Endogeneity Pitfall in Using Market Signals as Triggers for Prompt Corrective Action." Manuscript, University of Zurich.

Bond, Philip, Itay Goldstein, and Edward Simpson Prescott. 2010. "Market-Based Corrective Actions." *The Review of Financial Studies* 23 (February): 781–820.

Calomiris, Charles W., and Richard J. Herring. 2011. "Why and How to Design a Contingent Convertible Debt Requirement." Wharton Financial Institutions Center Working Paper 11-41 (April).

[13] Implicitly, these models assume that bank managers act in the best interest of equity owners. That assumption is, of course, debatable.

Credit Suisse Group. 2011. "Buffer Capital Notes Information Memorandum." Available at www.credit-suisse.com/investors/doc/buffer_capital_notes_information_memorandum.pdf.

Davis, Douglas, Oleg Korenok, and Edward Simpson Prescott. 2011. "An Experimental Analysis of Contingent Capital Triggering Mechanisms." Federal Reserve Bank of Richmond Working Paper 11-01R.

Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (May): 383–417.

Flannery, Mark J. 1998. "Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence." *Journal of Money, Credit and Banking* 30 (August): 273–305.

Flannery, Mark J. 2005. "No Pain, No Gain? Effecting Market Discipline via 'Reverse Convertible Debentures."' In *Capital Adequacy Beyond Basel: Banking, Securities, and Insurance,* edited by Hal S. Scott. Oxford: Oxford University Press, 171–96.

Flannery, Mark J. 2009. "Stabilizing Large Financial Institutions with Contingent Capital Certificates." Manuscript, University of Florida.

Government Accountability Office. 2011. "Bank Regulation: Modified Prompt Corrective Action Framework Would Improve Effectiveness." Washington, D.C.: GAO-11-612 (June).

Hart, Oliver, and Luigi Zingales. 2011. "A New Capital Regulation for Large Financial Institutions." *American Law and Economics Review* 13 (Fall): 453–90.

Huertas, Thomas F. 2009. "Too Big to Fail, Too Complex to Contemplate: What to do about Systemically Important Firms." Paper presented at the Financial Markets Group, London School of Economics, September 15.

Independent Commission on Banking. 2011. "Final Report Recommendations." Available at http://bankingcommission.s3.amazonaws.com/wp-content/uploads/2010/07/ICB-Final-Report.pdf.

Marshall, David A., and Edward Simpson Prescott. 2001. "Bank Capital Regulation with and without State-Contingent Penalties." *Carnegie-Rochester Conference Series on Public Policy* 54 (June): 139–84.

Marshall, David A., and Edward Simpson Prescott. 2006. "State-Contingent Bank Regulation with Unobserved Actions and Unobserved Characteristics." *Journal of Economic Dynamics and Control* 30 (November): 2,015–49.

McDonald, Robert L. 2011. "Contingent Capital with a Dual Price Trigger." Manuscript, Northwestern University.

Offices of Inspector General. 2010. "Evaluation of Federal Regulatory Oversight of Washington Mutual Bank." Report No. EVAL-10-002. Washington, D.C.: Federal Deposit Insurance Corporation (April).

Pennacchi, George. 2011. "A Structural Model of Contingent Bank Capital." Manuscript, University of Illinois.

Pennacchi, George, Theo Vermaelen, and Christian C. P. Wolff. 2010. "Contingent Capital: The Case for COERCs." INSEAD Working Paper 2010/55/FIN.

Plosser, Charles I. 2010. "Convertible Securities and Bankruptcy Reforms: Addressing Too Big to Fail and Reducing the Fragility of the Financial System." Paper presented at the Conference on the Squam Lake Report, New York, N.Y., June 16.

Prescott, Edward Simpson. 2001. "Regulating Bank Capital Structure to Control Risk." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (Summer): 35–52.

Raviv, Alon. 2004. "Bank Stability and Market Discipline: Debt-for-Equity Swaps versus Subordinated Notes." Available at http://129.3.20.41/eps/fin/papers/0408/0408003.pdf.

Roll, Richard. 1984. "Orange Juice and Weather." *American Economic Review* 74 (December): 861–80.

Squam Lake Group. 2010. *The Squam Lake Report: Fixing the Financial System*. Princeton, N.J.: Princeton University Press.

Stern, Gary H. 2001. "Taking Market Data Seriously." *The Region* Federal Reserve Bank of Minneapolis (September).

Sundaresan, Suresh, and Zhenyu Wang. 2011. "On the Design of Contingent Capital with Market Trigger." Federal Reserve Bank of New York Staff Report 448 (November).

White, Lawrence J. 1991. *The S&L Debacle*. New York: Oxford University Press.

Economic Quarterly—Volume 98, Number 1—First Quarter 2012—Pages 51–76

Exchange Rate Volatility in a Simple Model of Firm Entry and FDI

Thomas A. Lubik and Katheryn N. Russ

In recent years, the field of international trade has experienced a renaissance in theory and measurement, much of which is rooted in the seminal contribution by Melitz (2003). Melitz's theory of heterogeneous firms and entry has changed not only how the field understands trade flows, but also how it views multinational production. It enables more realistic modeling of multinational firm behavior by capturing the fact that only the largest and most efficient manufacturing firms invest abroad and, most importantly, that they earn positive profits.

In this article, we present and analyze a simple model of firm exit and entry in a Melitz-type environment. We apply the notion of endogenous variation in the entry margin to location decisions by domestic and foreign firms. If a firm wants to supply markets abroad, it has to locate production facilities in the foreign country. We interpret the outcome of this decision as foreign direct investment (FDI). Modeling this location decision thus links the theory of FDI with models of multinational enterprises (MNEs). Moreover, this has implications for the determination of international prices and quantities and related macroeconomic issues.

We want to accomplish two things with this article. First, we derive and explain a full set of analytical solutions for all variables of interest in our theoretical model. This comes at the price of some arguably restrictive assumptions. However, by doing so we can cleanly isolate the entry

Lubik is a senior economist and research advisor at the Richmond Fed. Russ is an assistant professor at the University of California, Davis. The authors are grateful to John Muth, Pierre Sarte, and Felipe Schwartzman, whose comments greatly improved the exposition of this article. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mails: thomas.lubik@rich.frb.org; knruss@ucdavis.edu.

mechanism that is at the core of the model and carries over to the variables in the model. Our contribution thus lies in making this mechanism more transparent compared to richer modeling environments that have to rely on numerical solutions. We therefore see this article as an introductory guide to the mechanics of Melitz-style models of multinational firms.

Second, we use the model to take a look at a perennial issue in international finance, namely the determinants of exchange rate volatility and the apparent disconnect with economic fundamentals. Recent discussions of exchange rate determination have increasingly emphasized the possible role of payments earned on FDI and other assets held abroad. Yet, there are few existing models of MNEs and endogenous exchange rates. This article demonstrates that the entry decisions of MNEs influence the volatility of the real exchange rate in countries where there are significant costs involved in maintaining production facilities, even when prices are perfectly flexible. We show that for plausible parameterizations, MNE activity can make the exchange rate more volatile than relative consumption.

The key element of this framework is that a firm's technology depends both on aggregate and idiosyncratic labor productivity. Given fixed costs of entry, this determines a firm-specific threshold productivity level, below which firms do not operate. This threshold moves around with aggregate economic conditions. Moreover, the model implies an endogenous distribution of firm-level productivities that has strong empirical support (e.g., Helpman, Melitz, and Yeaple 2004). In our model, the threshold or entry margin influences the relative volatility of exchange rates, the aggregate price level, and consumption arising in response to productivity shocks. A positive country-specific productivity shock allows both native and foreign-owned firms with lower firm-specific levels of productivity to become profitable. Lower idiosyncratic labor productivity in these new entrants dampens the impact of the country-specific shock on total aggregate productivity. Thus, a positive productivity shock can impact the real exchange rate at the same time entry by progressively less productive firms dampens the effect of the productivity shock on the aggregate price level and consumption.

In order to highlight the entry channel for exchange rate determination and to derive closed-form solutions, we make two simplifying assumptions. First, we segment markets by allowing no cross-border transfers of wealth via portfolio investment and we shut down any real trade linkages, except for those involving the production and remittance activities of multinational firms. These assumptions leave the nominal exchange rate completely determined by flows of currency used for paying local costs of production incurred by overseas branches of MNEs and for repatriating their profits earned abroad. We show that FDI, even in this model without sunk costs of physical capital, can act as the key driver of real and nominal exchange rate movements.

The second assumption is standard in the Melitz-type literature, namely that participation in any market is a period-by-period decision. This simplifies the model's solution considerably since it eliminates the presence of endogenous state variables in the solution. In addition, it yields testable empirical implications linking a country's industrial structure to the volatility of its exchange rate. We find that the behavior of multinationals is most likely to generate excess volatility when FDI is plentiful in sectors with higher industry concentration, higher value-added, and higher barriers to foreign participation relative to domestic production, so that foreign firms tend to be big relative to domestic firms.

The rest of the article considers the role that MNEs can play in explaining the determinants of exchange rates. We begin by placing our analysis within the broader context of the recent literature. We then introduce a simple, stylized model of multinational production. We emphasize the role of entry in determining the aggregate productivity level and the number of different goods available in the economy. Section 3 contains the main analysis of the model. We discuss intuitively the role that market entry plays in the response to shocks to technology for both nominal and real exchange rates, as well as for consumption and other real quantities. We show analytically how this can be decomposed into direct and indirect effects. We then discuss the implication of our model for the exchange rate disconnect puzzle and the volatility puzzle. The last section concludes.

1. RELATION TO THE LITERATURE

It is well known that the volatility of the exchange rate is much higher than that of other macroeconomic variables, such as the aggregate price level and consumption. This produces a fundamental challenge for optimization-based open economy models that link marginal rates of substitution to international goods prices. For instance, Baxter and Stockman (1989) and Flood and Rose (1995) point out that nominal and real exchange rate volatility is typically 10 times higher than the volatility of relative prices and several times greater than the volatility of output or consumption. As demonstrated by Backus, Kehoe, and Kydland (1992), standard open economy business cycle models have difficulty replicating these stylized facts unless implausible substitution elasticities are assumed. The reason is the tight link between marginal rates of substitution and international relative prices that are at the heart of optimization-based frameworks.

This exchange rate volatility puzzle is related to, in the nomenclature of Rogoff (1996), the exchange rate disconnect puzzle. It stipulates that, empirically, exchange rates appear to behave virtually independently of underlying economic fundamentals. Consequently, the ability of modern open economy macroeconomics to explain exchange rate movements has not been

an unqualified success.[1] In this article, we approach this issue not from the goods side, but rather from a perspective of financial flows generated by the operations of MNEs. This removes the burden of having relative quantities match the volatility of relative prices.[2] In order to capture in the model the disconnect between relative consumption and international prices, we turn to the literature on MNEs and FDI, which de-emphasizes the role of final consumption in favor of production decisions.

Our model draws its motivation from this growing body of work that stresses the potential role of MNEs as one factor driving exchange rate fluctuations. We add the additional consideration that entry by heterogeneous firms affects fluctuations in prices and consumption, and thus exchange rate volatility. Quantitatively, there are several studies that highlight a causal relationship between FDI and the exchange rate. Kosteletou and Liargovas (2000) provide empirical evidence that inflows of FDI Granger-cause fluctuations in the real exchange rate for some European countries. Whether FDI generates appreciating or depreciating tendencies varies by country, a disparity that the authors explain as emerging from each country's use of the inflows to finance either consumption or capital accumulation. Shrikhande (2002) builds a theoretical model that allows for cross-border acquisitions of physical capital. He is able to replicate the observed persistence and time-varying volatility in the real exchange rate using fixed investment costs, similar to the fixed cost of entry in our model. Gourinchas and Rey (2007) find empirical evidence of a recursive relationship between exchange rates and the return on net foreign asset holdings, including FDI, such as we model here.

Whereas the reduced-form correlation between FDI and exchange rate volatility is well established, the direction of causality is widely debated. Specifically, the literature seeking to measure the effect of exchange rate volatility on FDI is vast and conflicted, which further supports the analysis in this article linking them both as endogenous variables. Phillips and Ahmadi-Esfahani (2008) provide an exhaustive survey of these varied empirical and theoretical results. Several articles have recently analyzed entry and production behavior of heterogeneous multinational firms. Russ (2007, 2011) shows that accounting for the source of exchange rate volatility can determine whether the relationship between volatility and FDI is positive or negative. Fillat and Garetto (2010) find evidence that increased uncertainty of any type in the host country can increase the likelihood that firms will export rather than invest abroad. Ramondo, Rappoport, and Ruhl (2010) obtain the result

[1] The seminal article in this literature is Meese and Rogoff (1983). Different perspectives on this issue are given by Clarida and Galí (1994) in a value-at-risk framework, and Lubik and Schorfheide (2005) in an estimated dynamic stochastic general equilibrium model.

[2] We should point out, however, that we do not speak to the other part of the disconnect puzzle, namely that exchange rates are essentially unpredictable. This issue is left for a much more empirical treatment than the scope of this article allows.

that real exchange rate volatility can be correlated with lower multinational production relative to arms-length exports when real wages and employment are fixed.

There are important conceptual, empirical, and purely practical reasons for modeling multinational firms characterized by heterogeneous productivity levels. First, it is difficult to explain why some firms, but not all, establish branches abroad, unless there exists some differential in their potential to make a profit, as would occur when firms have differing labor productivity. Second, there are several stylized facts regarding the behavior of MNEs that conflict with the representative firm assumption. Using an extensive data set that joins observations on firm size and employment with intra- and inter-firm trade data, Bernard, Jensen, and Schott (2009) show that multinational firms are larger in size and have greater revenues per worker than firms that do not show evidence of having overseas affiliates. Modeling firm-specific labor productivity as Pareto-distributed generates a pattern of firm sizes that is also Pareto, which conforms to empirical findings by Helpman, Melitz, and Yeaple (2004) and di Giovanni, Levchenko, and Rancière (2011), among others. These stylized facts of firm size and distribution are captured by the heterogeneous firm framework.

Finally, introducing heterogeneity in the tradition of Melitz (2003) causes the entire solution of the model to rest only on the lowest productivity level among firms producing in a particular period and a set of exogenous parameters. Pinpointing this threshold productivity level using a zero-profit cutoff condition allows the entire model to be solved numerically without linearization and yields analytical results depicting the influence of shocks to a country's general technological state on the nominal and real exchange rate.

The mechanism we identify, namely that aggregate consumption and prices appear to be much less volatile than the exchange rate because their movement in response to a positive country-specific productivity shock can be dampened by the entry of less productive domestic firms, is akin to a new vein of literature on the exchange rate disconnect puzzle emphasizing the role of transaction costs in trade. Fitzgerald (2008) shows both theoretically and empirically that trade costs based on the geographic distance between countries can explain why relative price levels are much less volatile than the real exchange rate, even when prices are perfectly flexible. Our article abstracts from trade in goods, all local consumption being produced by either domestic firms or resident branches of MNEs. It nonetheless approaches the disconnect puzzle in a similar spirit, asking not why nominal and real exchange rates are so volatile, but why they appear so volatile relative to consumption and relative price levels.

The model closest to ours is Cavallari (2007), which demonstrates that in a framework with heterogeneous firms, exchange rate overshooting may be generated by repatriated profits from multinational firms exploiting a positive

productivity shock overseas. Cavallari relies on sticky prices to drive the result. We show, on the other hand, that entry behavior alone can create exchange rate volatility exceeding that of fundamentals, even with flexible prices. As opposed to the model in Ghironi and Melitz (2005), our framework does not involve the sunk costs or incomplete asset markets that generate, respectively, endogenous persistence in exchange rate behavior and a role for active monetary policy in a study of heterogeneous exporters and exchange rates. However, it is rich enough to demonstrate that production decisions by multinational firms can explain part of the differential in the variance of exchange rates and other macroeconomic variables without nominal rigidities.

2. A SIMPLE MODEL OF ENTRY AND FDI

Our model economy consists of two countries, Home and Foreign, that are identical in every respect. Each country is composed of a representative consumer and a continuum of firms. The consumer enjoys the consumption of goods supplied by both Home and Foreign firms, but derives disutility from supplying labor to firms operating in his home country. Home and Foreign firms are distributed along separate unit intervals. What classifies a firm as Foreign is that it pays a fixed overhead cost denominated in the currency of its host country and repatriates (nominal) profits earned at the end of each production period. This creates a necessity for foreign exchange since the firm's owners can only buy goods using their own home currency.

Furthermore, we assume that there is no trade in goods. Foreign firms can supply the domestic market only by opening production facilities there. Consequently, there is no trade balance, only a capital account in the balance of payments. We also abstract from international borrowing and lending. However, consumers can hold financial wealth in the form of currency that is issued by each country's monetary authority. The relative supply of the two currencies is one of the determinants of the nominal exchange rate.

The Consumer's Problem

The representative consumer in the Home country maximizes lifetime utility

$$\max_{\{C_t, L_t, M_{t+1}\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t U(C_t, L_t) \right], \tag{1}$$

subject to the budget constraint

$$P_t C_t + M_{t+1} \leq W_t L_t + M_t + \Pi_t + T_t, \tag{2}$$

and the cash-in-advance constraint

$$P_t C_t \leq M_t. \tag{3}$$

C_t is aggregate consumption, L_t is labor input, M_t is the money stock, W_t is the nominal wage, Π_t are firm profits accruing to the household, and T_t are transfer payments from the government; P_t is the aggregate price index, which we define below. The household discounts future utility streams with $0 < \beta < 1$.

We assume that the period utility function is additively separable, $U(C_t, L_t) = \frac{C_t^{1-\rho}-1}{1-\rho} - \chi L_t$, where $\rho > 0$, $\chi > 0$. Furthermore, we define the consumption aggregator as

$$C_t = \left[\int_0^{n_{h,t}} c_{h,t}(i)^{\frac{\theta-1}{\theta}} di + \int_1^{1+n_{f,t}} c_{f,t}(i)^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}}, \tag{4}$$

with $\theta > 1$. The interval $[0, n_{h,t})$ represents the continuum of all goods $c_{h,t}(i)$ that can possibly be produced by Home-owned firms for the Home market, while the interval $[1, 1 + n_{f,t}]$ represents the continuum of all goods that can be produced by Foreign-owned firms, $c_{f,t}(i)$, for the Home market $(n_{h,t}, n_{f,t} \leq 1)$. The specification of the sub-utility function C_t encapsulates a preference for variety in that it is increasing in the number of firms $n_{h,t} + n_{f,t}$ supplying the market. Variations in this extensive margin through entry will therefore be another determinant of the exchange rate.

In solving our model, we assume that the cash-in-advance constraint always binds. This determines aggregate consumption as a function of real money balances, $C_t = \frac{M_t}{P_t}$. From the consumption aggregator, we can derive demand equations for individual goods produced by Home and Foreign firms that are downward sloping in relative prices. Homothetic preferences imply that the demand for each good is a constant proportion of aggregate consumption:

$$\begin{aligned} c_{h,t}(i) &= \left(\frac{p_{h,t}(i)}{P_t} \right)^{-\theta} C_t, \\ c_{f,t}(i) &= \left(\frac{p_{f,t}(i)}{P_t} \right)^{-\theta} C_t. \end{aligned} \tag{5}$$

Finally, the optimality condition for total labor input yields a wage equation:

$$W_t = \chi P_t C_t^{\rho}. \tag{6}$$

The Firm's Problem

In each country, there is a continuum of firms with plans to put their particular invention into production. We denote firms owned by residents of the Home country with the label h, while firms owned by residents of the Foreign country carry f. The location of production is identified by a "$*$" for Foreign, and no label for Home. Every firm that decides to enter the market

during period t produces a unique good and operates under a unique, firm-specific productivity level, $\varphi(i)$. We assume that idiosyncratic productivity has a continuous distribution $g(\varphi)$, with support over the interval $(0, \infty)$. Any difference among the pricing rules and production decisions of firms operating in the Home country is due only to differences in φ. Thus, φ is used to index each good and the firm that produces it, instead of the general subscript i.

This idiosyncratic component is distinct from an aggregate time-varying disturbance A_t, which denotes the country-specific state of technology available to all firms operating in the Home country. Technology is thus characterized by

$$y_{h,t}(i) = A_t \varphi(i) l_{h,t}(i), \tag{7}$$

where $l_{h,t}(i)$ is the amount of labor used by Home firm i for production in the Home country. The country-specific productivity parameter for the Home country, A_t, is defined by

$$\log A_t = \phi \log A_{t-1} + \varepsilon_{At},$$

where $\varepsilon_{At} \sim N(0, \sigma^2_{\varepsilon_A})$.

Home firms operating in the Home country maximize profits subject to consumer demand. They also bear a fixed overhead cost of production, f_h, denominated in units of aggregate output. The profit maximization problem is thus

$$\max_{p_{h,t}(\varphi)} \left\{ \pi_{h,t}(\varphi) = p_{h,t}(\varphi) c_{h,t}(\varphi) - W_t \frac{c_{h,t}(\varphi)}{\varphi A_t} - P_t f_h \right\}$$

$$\text{s.t.} \quad c_{h,t}(\varphi) \leq \left(\frac{p_{h,t}(\varphi)}{P_t} \right)^{-\theta} C_t, \tag{8}$$

where we have used the market clearing condition $c_{h,t}(\varphi) = y_{h,t}(\varphi)$, and substituted out labor input $l_{h,t}(\varphi)$ with the production function. Assuming an interior solution, that is, where the firm has already entered and operates in the Home market, the first-order condition for profit maximization is then

$$\frac{\partial \pi_{h,t}(\varphi)}{\partial p_{h,t}(\varphi)} : c_{h,t}(\varphi) + \frac{\partial c_{h,t}(\varphi)}{\partial p_{h,t}(\varphi)} p_{h,t}(\varphi) - \frac{\partial c_{h,t}(\varphi)}{\partial p_{h,t}(\varphi)} \frac{W_t}{\varphi A_t} = 0. \tag{9}$$

We can now derive the optimal price-setting condition by substituting the derivative of the demand equation into the firm's first-order condition:

$$p_{h,t}(\varphi) = \frac{\theta}{\theta - 1} \frac{W_t}{\varphi A_t}. \tag{10}$$

As is typical in models with Dixit-Stiglitz-type preferences for variety, firms set prices as a markup over marginal costs. Moreover, for a given wage, higher productivity firms charge a lower price since they have lower marginal costs.

The same steps can be used to derive the pricing equation for Foreign-owned firms operating in the Home country. The profit maximization problem

is

$$\max_{p_{f,t}(\varphi)} \left\{ \pi_{f,t}(\varphi) = \left(\frac{1}{S_t}\right) \left[p_{f,t}(\varphi) c_{f,t}(\varphi) - W_t \frac{c_{f,t}(\varphi)}{\varphi A_t} - P_t f_f \right] \right\}$$
$$\text{s.t. } c_{f,t}(\varphi) \leq \left(\frac{p_{f,t}(\varphi)}{P_t}\right)^{-\theta} C_t, \tag{11}$$

where S_t is the nominal exchange rate at time t, measured in units of Home currency per unit of Foreign currency. The term f_f denotes the fixed cost paid by Foreign-owned firms operating in the Home country. The fixed cost is denominated in units of the aggregate output of the host country and paid in units of local currency. It can be thought of as an overhead cost, or, more abstractly, as the cost of capital with 100 percent depreciation. The pricing rule for Foreign goods produced and sold in the Home country turns out to be identical, since firms face the same Home-country wage and are influenced by the same country-specific productivity shocks:

$$\frac{\partial \pi_{f,t}(\varphi)}{\partial p_{f,t}(\varphi)} : \left(\frac{1}{S_t}\right) c_{f,t}(\varphi) + \left(\frac{1}{S_t}\right) \frac{\partial c_{f,t}(\varphi)}{\partial p_{f,t}(\varphi)} p_{f,t}(\varphi) - \left(\frac{1}{S_t}\right) \frac{\partial c_{f,t}(\varphi)}{\partial p_{f,t}(\varphi)} \frac{W_t}{\varphi A_t} = 0, \tag{12}$$

from which it follows immediately that

$$p_{f,t}(\varphi) = \frac{\theta}{\theta - 1} \frac{W_t}{\varphi A_t}. \tag{13}$$

More productive firms, that is, those having a high level of labor productivity φ, will charge lower prices, sell more units, and earn higher revenues and profits.

We now define a few more concepts that will prove useful in solving the model. Let $\eta_{h,t}(\varphi)$ and $\eta_{f,t}(\varphi)$ be the distributions of firm-specific productivity levels observed among active Home- and Foreign-owned firms. The aggregate price level P_t, which is the price index that minimizes expenditure on a given quantity of the aggregate consumption index in equation (4) is then given by[3]

$$P_t = \left[n_{h,t} \int_0^\infty p_{h,t}(\varphi)^{1-\theta} \eta_{h,t}(\varphi) d\varphi + n_{f,t} \int_0^\infty p_{f,t}(\varphi)^{1-\theta} \eta_{f,t}(\varphi) d\varphi \right]^{\frac{1}{1-\theta}}. \tag{14}$$

Substituting the pricing rules for individual goods, the expression reduces to

$$P_t = \left(\frac{\theta}{\theta - 1}\right) \frac{W_t}{A_t} \left[n_{h,t} \int_0^\infty \varphi^{\theta-1} \eta_{h,t}(\varphi) d\varphi + n_{f,t} \int_0^\infty \varphi^{\theta-1} \eta_{f,t}(\varphi) d\varphi \right]^{\frac{1}{1-\theta}}. \tag{15}$$

[3] See Melitz (2003) and Russ (2007) for a discussion of the computation of the aggregate price level and average firm-specific level of labor productivity.

We now define the average firm-specific productivity level for firms owned by country j as $\bar{\varphi}_{j,t} = \left[\int_0^{\infty} \varphi^{\theta-1} \eta_{j,t}(\varphi) d\varphi\right]^{\frac{1}{\theta-1}}$. It follows that the production-weighted average firm-specific level of labor productivity $\bar{\varphi}_t$ can be written as

$$\bar{\varphi}_t = \left[\frac{n_{h,t}}{N_t}\bar{\varphi}_{h,t}^{\theta-1} + \frac{n_{f,t}}{N_t}\bar{\varphi}_{f,t}^{\theta-1}\right]^{\frac{1}{\theta-1}}, \tag{16}$$

where $N_t = n_{h,t} + n_{f,t}$ is the composite continuum of goods available in the Home economy, which equals the number of firms. Using these expressions for average firm productivity together with the wage equation (6) and the cash-in-advance constraint in (15), we can finally express the aggregate price level as

$$P_t = \left(\chi \frac{\theta}{\theta - 1} \frac{N_t^{\frac{1}{1-\theta}}}{\bar{\varphi}_t A_t}\right)^{\frac{1}{\rho}} M_t. \tag{17}$$

The price level is decreasing in the number of goods available since consumers have a preference for variety, which makes for a more expensive consumption bundle. It is decreasing in aggregate productivity and the index of average idiosyncratic productivities.

The Zero-Profit Cutoff Condition

The production side of the economy is characterized by a continuum of prospective Home and Foreign entrepreneurs distributed, respectively, over $[0, 1)$ and $[1, 2]$, but only firms that can expect to be sufficiently productive to recoup the overhead cost will choose to produce in a particular period. Any firm may enter, depending on whether its total productivity, φA_t, is high enough to result in revenues sufficient to cover this per-period fixed cost.

We now determine the idiosyncratic productivity level that is sufficient for a firm to generate non-negative revenue net of entry costs. We identify the lowest productivity level, $\hat{\varphi}$, that allows a firm to enter into production using the Zero-Profit Cutoff (ZPC) condition. Formally, the ZPCs for Home- and Foreign-owned firms operating in the Home country are given by

$$\pi_{h,t}(\hat{\varphi}_{h,t}) = p_{h,t}(\hat{\varphi}_{h,t}) c_{h,t}(\hat{\varphi}_{h,t}) - W_t l_{h,t}(\hat{\varphi}_{h,t}) - P_t f_h \stackrel{!}{=} 0 \tag{18}$$

and

$$\pi_{f,t}(\hat{\varphi}_{f,t}) = \left(\frac{1}{S_t}\right)\left[p_{f,t}(\hat{\varphi}_{f,t}) c_{f,t}(\hat{\varphi}_{f,t}) - W_t l_{f,t}(\hat{\varphi}_{f,t}) - P_t f_f\right] \stackrel{!}{=} 0, \tag{19}$$

respectively. Analogous expressions apply to entry in the Foreign market.

We substitute the optimal pricing equation, the goods demand function, and the expression for real balances into the respective ZPCs. After straightforward, but tedious algebra, we arrive at the following intermediate expression for the productivity threshold values:

$$\hat{\varphi}_{j,t} = \theta f_j \left(\frac{\theta}{\theta - 1} \frac{\chi}{A_t} \right)^{1/\rho} \left(\bar{\varphi}_t N_t^{\frac{1}{\theta-1}} \right)^{\theta-1-1/\rho}, \quad j = h, f. \tag{20}$$

The threshold values are identical for both Home and Foreign firms except for the differences in the fixed cost of entry. Furthermore, the difference between the thresholds depends only on the ratio of the fixed costs they pay to produce in the Home market:

$$\hat{\varphi}_{f,t} = \left(\frac{f_f}{f_h} \right)^{\frac{1}{(\theta-1)}} \hat{\varphi}_{h,t}. \tag{21}$$

Firms with a higher entry cost need to have higher productivity to stay active. This is a recurring theme in the FDI literature, as there is substantial empirical evidence showing that only the highest-productivity firms engage in foreign direct investment.

We can derive similar expressions from the ZPCs for the Foreign country:

$$\hat{\varphi}^*_{j,t} = \theta f^*_j \left(\frac{\theta}{\theta - 1} \frac{\chi}{A^*_t} \right)^{1/\rho} \left(\bar{\varphi}^*_t N_t^{*\frac{1}{\theta-1}} \right)^{\theta-1-1/\rho}, \quad j = h, f. \tag{22}$$

The structure of the threshold condition is identical to the one for the Home country, but we allow for potentially different entry costs. Moreover, we assume that the fixed cost involved in production abroad is sufficiently large that a firm producing abroad will always produce in its native country as well ($\hat{\varphi}^*_{f,t} \leq \hat{\varphi}_{f,t}$). Thus, our model does not capture issues of geographic preference in firm location.

There are two *ceteris paribus* observations we can make at this stage. First, the threshold is decreasing in aggregate productivity. In a cyclical upswing, firms' idiosyncratic productivity does not have to be as high to generate enough revenue to cover the fixed cost. Second, for large enough values of the substitution elasticity θ, the threshold is increasing in both the average productivity $\bar{\varphi}_t$ and the number of firms operating in the home country N_t. Both effects reflect the influence of competition. The marginal firm operating in the Home country needs to have higher idiosyncratic productivity, both when average firm productivity is higher and when it is competing with a large number of other firms. We should note, however, that these deliberations are partial equilibrium in nature, as an exogenous rise in aggregate productivity presumably increases the number of firms, but lowers average productivity. In order to go much further we now need to make distributional assumptions on the nature of the firm-specific productivity process. Once $g(\varphi)$ is specified, equation (20) is sufficient to pinpoint the minimum level of labor productivity for Home and Foreign firms entering the Home market.

The Number of Firms

As described in Helpman, Melitz, and Yeaple (2004) and Russ (2007), the equilibrium distribution of firm-specific productivity levels for firms owned by country $j \in [h, f]$ is truncated, so that firms with productivity levels too low to earn at least zero profits do not produce in period t. These low-productivity firms are plucked from the formulation of the aggregate price and output levels, leaving a truncated equilibrium distribution:

$$\eta_{j,t}(\varphi) = \left\{ \begin{array}{c} 0 \text{ for } \varphi < \hat{\varphi}_{j,t} \\ \frac{g(\varphi)}{1-G(\hat{\varphi}_{j,t})} \text{ for } \varphi > \hat{\varphi}_{j,t} \end{array} \right\}. \tag{23}$$

This allows us to determine the number of firms in the economy. Denote $n_{j,t}$ for firms owned by residents of country j who enter the Home market ($j \in [h, f]$). It follows that this is simply the probability that *any* firm holds an idiosyncratic productivity parameter greater than $\hat{\varphi}_{j,t}$. Specifically, $n_{j,t} = 1 - G(\hat{\varphi}_{j,t})$. For instance, as $\hat{\varphi}_{f,t}$ increases, the proportion of Foreign-owned firms entering the Home market falls. Such an increase means that a Foreign firm must have a greater idiosyncratic level of labor productivity to expect to enter without incurring a loss. We can thus write average productivity levels in the Home- and Foreign-owned sector as

$$\bar{\varphi}_{j,t} = \left[\frac{1}{1 - G(\hat{\varphi}_{j,t})} \int_{\hat{\varphi}_{j,t}}^{\infty} \varphi^{\theta-1} g(\varphi) d\varphi \right]^{\frac{1}{\theta-1}}, \; j = h, f. \tag{24}$$

Using this expression and the definition of productivity index (16) we find that

$$\bar{\varphi}_t N_t^{\frac{1}{\theta-1}} = \left[\int_{\hat{\varphi}_{h,t}}^{\infty} \varphi^{\theta-1} g(\varphi) d\varphi + \int_{\hat{\varphi}_{f,t}}^{\infty} \varphi^{\theta-1} g(\varphi) d\varphi \right]^{\frac{1}{\theta-1}}. \tag{25}$$

We further assume, for purposes of exposition, that idiosyncratic productivity is drawn from a Pareto distribution. The Pareto distribution is used widely in the literature on firm entry and FDI as it describes firm size and rank distribution well. Specifically, the probability and cumulative density functions are given by, respectively, $g(\varphi) = k\varphi^{-(k+1)}$ and $G(\varphi) = 1 - \varphi^{-k}$, with the shape parameter $k > 0$.[4] This specification now allows us to compute the integrals in the above expression. After several steps, using condition (21), we can solve for the threshold productivity level for Home firms as a function of the exogenous aggregate productivity shock alone:

$$\hat{\varphi}_{h,t} = \psi_0 A_t^{-\frac{\theta-1}{k\rho(\theta-1)+(\theta-1)-k}}, \tag{26}$$

[4] We normalize the location parameter of the distribution to 1, which implies a support of $[1, \infty)$.

where ψ_0 is a constant.[5] We note that for the underlying Pareto distribution to have bounded variance, we need $k > 2$. Furthermore, for the integral to exist, we have to assume $k > \theta - 1$. We also note that the distribution of firm-specific productivity induces a distribution over the idiosyncratic productivities of active firms in the Home country, which is Pareto itself.

It can be easily verified that the exponent on A_t in equation (26) is positive for all ranges of parameter values for ρ and θ. However, if the Pareto shape parameter k becomes very large relative to the coefficient of relative risk aversion, then firms become less disperse (that is, heterogeneity becomes less important). Moreover, θ should not be too large relative to k since otherwise consumers' love of variety is not strong enough to keep them from buying only the cheapest goods. We thus find that an increase in aggregate productivity leads to a fall in the threshold level of idiosyncratic productivity that firms need to cross in order to cover the fixed costs of operation.

We can now compute the remaining endogenous variables. The total number of varieties equals

$$N_t = n_{h,t} + n_{f,t} = \hat{\varphi}_{h,t}^{-k} + \hat{\varphi}_{f,t}^{-k} = \left[1 + \left(\frac{f_f}{f_h}\right)^{\frac{1}{(\theta-1)}}\right] \hat{\varphi}_{h,t}^{-k}. \tag{27}$$

An increase in aggregate productivity lowers the threshold for both Home- and Foreign-owned firms, which raises their numbers in the Home economy. It can also be quickly verified that the average firm-specific productivity level is

$$\bar{\varphi}_{j,t} = \left(\frac{k}{k - (\theta - 1)}\right)^{\frac{1}{\theta-1}} \hat{\varphi}_{j,t}, \;\; j = h, f.$$

It is proportional and increasing in the threshold level. An increase in $\hat{\varphi}_{j,t}$ reflects the exit of less productive firms, and thus implies that average idiosyncratic productivity rises. Similarly, the measure of aggregate firm productivity,

$$\bar{\varphi}_t = \left[1 + \left(\frac{f_h}{f_f}\right)^{\frac{k-(\theta-1)}{(\theta-1)}}\right]^{\frac{1}{\theta-1}} \left[1 + \left(\frac{f_f}{f_h}\right)^{\frac{1}{(\theta-1)}}\right]^{\frac{1}{1-\theta}} \left(\frac{k}{k - (\theta - 1)}\right)^{\frac{1}{\theta-1}} \hat{\varphi}_{h,t}, \tag{28}$$

is increasing in the Home- and Foreign-owned average productivities and thus the threshold, $\hat{\varphi}_{h,t}$.

---

[5] Specifically,

$$\psi_0 = \left\{ (\theta f_h)^{\rho} \left(\frac{\theta - 1}{\theta}\right)^{\kappa} \left[1 + \left(\frac{f_h}{f_f}\right)^{\frac{\kappa-(\theta-1)}{(\theta-1)}}\right]^{\frac{\rho(\theta-1)-1}{(\theta-1)}} \left(\frac{k}{k - (\theta - 1)}\right)^{\frac{\rho(\theta-1)-1}{(\theta-1)}} \right\}^{\frac{\theta-1}{k\rho(\theta-1)+(\theta-1)-k}}.$$

The Balance of Payments and the Exchange Rate

We now close the model by describing international transactions. There is no international borrowing and lending and domestic agents are restricted to holding only domestic currency. Moreover, there is no international trade in goods. Instead, domestic consumers can satisfy their demand for Foreign products from Foreign firms that have located their production facilities in the Home country. The only exchange across borders is through the repatriation of profits, as we assume that entry costs of Foreign firms have to be paid in terms of the host country's currency.

Equilibrium in the Foreign exchange market requires that the number of units of Home currency being offered for exchange by overseas branches of Foreign multinationals repatriating their profits equal the number of units of Home currency demanded by overseas branches of Home multinationals repatriating their own profits. This condition is the multinational analog to the condition for a world with exporters described in Bacchetta and van Wincoop (2000):[6]

$$S_t n^*_{h,t} \pi^*_{h,t}(\bar{\varphi}^*_{h,t}) = n_{f,t} \pi_{f,t}(\bar{\varphi}_{f,t}). \tag{29}$$

Using the ZPC condition and the solution for the average firm-specific productivity level, we have

$$\begin{aligned} n_{f,t}\pi_{f,t}(\bar{\varphi}_{f,t}) &= n_{f,t}\left[p_{f,t}(\bar{\varphi}_{f,t})c_{f,t}(\bar{\varphi}_{f,t}) - W_t l_{f,t}(\bar{\varphi}_{f,t}) - P_t f_f\right] \\ &= n_{f,t} P_t f_f \left[\left(\frac{k}{k-(\theta-1)}\right) - 1\right]. \end{aligned}$$

Applying the same process to the left-hand side of the balance-of-payments equation yields an expression for the nominal exchange rate:[7]

$$S_t = \frac{f_f}{f^*_h} \frac{n_{f,t}}{n^*_{h,t}} \frac{P_t}{P^*_t}. \tag{30}$$

The exchange rate is determined by three factors: first, the relative size of the entry costs; second, the number of firms operating in the respective foreign markets. This determines the overall volume of capital account transactions. *Ceteris paribus*, if the number of Foreign firms operating in the Home country is relatively large, then their domestic currency denominated profits have to be

[6] See Russ (2007) for a derivation of the aggregation of profits, which is also described in Melitz (2003). Intuitively, we have to aggregate over all individual Foreign-owned firms operating in the respective host countries. As it turns out, this can be expressed as the product of the profit of the firm with average productivity $\pi_{f,t}(\bar{\varphi}_{f,t})$ and the number of firms $n_{f,t}$. Similar reasoning applies to Home firms operating abroad.

[7] The expression is considerably simplified by the assumption that both countries are identical except for the exogenous shock processes. We regard this as a clean benchmark and a starting point for further work.

exchanged against a relatively smaller supply of foreign currency denominated profits. Hence, their relative value and thus the price of domestic currency is low (i.e., the exchange rate S_t is high). The third factor are domestic price levels, as in any quantity-theoretic model. What differentiates our framework from a standard exchange rate model is the presence of frictions in the form of entry costs into foreign markets.

We now perform the final steps in deriving an analytical solution for the model. The price level P_t in equation (17) depends on the total number of firms N_t, aggregate firm productivity $\bar{\varphi}_t$, the money supply M_t, and the exogenous shock A_t. We can substitute the reduced-form expressions for the endogenous variables in the price level equation, which yields

$$P_t = \psi_0^{\frac{k-(\theta-1)}{\rho(\theta-1)}} \psi_1 \left(\frac{1}{A_t}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}} M_t, \tag{31}$$

where $\psi_1 = \left(\frac{\kappa\theta}{\theta-1}\right)^{\frac{1}{\rho}} \left[1 + \left(\frac{f_f}{f_h}\right)^{\frac{1}{(\theta-1)}}\right]^{-\frac{1}{\rho(\theta-1)}} \left(\frac{k}{k-(\theta-1)}\right)^{-\frac{1}{\rho(\theta-1)}}$. The nominal price level is increasing in the money stock with unit elasticity, while it is decreasing in the productivity shock. We will make a quantitative assessment of the productivity elasticity below. We also note that the solution for aggregate consumption can be found from this expression, using the cash-in-advance constraint, that is, $C_t = \frac{M_t}{P_t}$.

We assume that both countries are identical with respect to their economic structure, except that they are driven by independent shocks. We can therefore write the price level ratio $\frac{P_t}{P_t^*} = \left(\frac{A_t^*}{A_t}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}} \frac{M_t}{M_t^*}$. These two expressions reflect the cash-in-advance constraint for money holding, which delivers a quantity-theoretic result, but with a twist. The relative and absolute price level is unit-elastic in money supply, but moves inversely with (relative) productivity. We also find it useful to compute the expression for the relative consumption ratios between the two countries, namely $\frac{C_t}{C_t^*} = \left(\frac{A_t}{A_t^*}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}}$.

We can now define the real exchange rate $RER_t = S_t \frac{P_t^*}{P_t} = \frac{f_f}{f_h^*} \frac{n_{f,t}}{n_{h,t}^*}$. In order to provide a closed-form solution, we need to determine the relative number of Foreign firms operating in their respective host countries, $\frac{n_{f,t}}{n_{h,t}^*}$. We note that $n_{f,t} = \hat{\varphi}_{f,t}$, namely the value of the productivity threshold. We can substitute this into the definition of the real exchange rate:

$$RER_t = \left(\frac{f_f}{f_h}\right) \left(\frac{A_t}{A_t^*}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}}. \tag{32}$$

The real exchange rate depends only on relative productivity levels. Since $\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k} > 0$, an increase in productivity at home increases the real exchange rate and the relative price of the domestic consumption bundle falls. This is the standard supply effect on the real exchange rate, as ceteris paribus

the productivity increase leads to higher output, lower prices, and thus a lower price level, which makes Foreign-produced goods more expensive. The elasticity coefficient is the same as the one we identified before in the price level. This shows that real exchange rate movements are driven by real factors. This conjecture is borne out when we compute the nominal exchange rate:

$$S_t = \frac{f_f}{f_h^*}\frac{n_{f,t}}{n_{h,t}^*}\frac{P_t}{P_t^*} = \frac{f_f}{f_h}\frac{M_t}{M_t^*}. \tag{33}$$

In the absence of any nominal friction, there is no effect of the money supply on real variables.

Closing the Model

We now discuss the remaining general equilibrium and aggregation conditions that close the model. Expressions for all reduced-form solutions are listed in Table 1. We first compute the solution for the labor supply. Noting that $c_{h,t}(i) = y_{h,t}(i)$, we can use the firm-specific demand function in equation (5) and the production function in equation (7) to find labor input for firm i:

$$\begin{aligned} l_{h,t}(i) &= \left(\frac{p_{h,t}(i)}{P_t}\right)^{-\theta}\frac{C_t}{A_t\varphi\,(i)} = \\ &= \left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta}A_t^{\theta-1}\varphi\,(i)^{\theta-1}. \end{aligned} \tag{34}$$

The second line is derived by using the solution for firm i's optimal price (10) and the wage (6). This relationship applies to all firms making non-negative profits.

We can thus aggregate over all Home firms that operate domestically:

$$\begin{aligned} L_{h,t} &= \int_0^{n_{h,t}} l_{h,t}(i)di = \left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta}A_t^{\theta-1}\int_0^{n_{h,t}}\varphi\,(i)^{\theta-1}\,di \\ &= \frac{1}{\theta}\left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta}A_t^{\theta-1}\hat{\varphi}_{h,t}^{-k(\theta-1)}, \end{aligned} \tag{35}$$

where the last equality uses $n_{h,t} = \hat{\varphi}_{h,t}^{-k}$. We can derive a virtually identical expression for Foreign firms operating in the Home market, whereby we rely on the assumption that they face the same demand schedules and the same labor market. The only difference is that Foreign firms pay a higher fixed cost for entry, which results in a higher productivity threshold.

Table 1 Closed-Form Solutions

Variables

$\hat{\varphi}_{h,t} = \psi_0 A_t^{-\frac{\theta-1}{k\rho(\theta-1)+(\theta-1)-k}}$	Productivity Threshold Home Firms
$\hat{\varphi}_{f,t} = \left(\frac{f_f}{f_h}\right)^{\frac{1}{(\theta-1)}} \hat{\varphi}_{h,t}$	Productivity Threshold Foreign Firms
$n_{h,t} = \hat{\varphi}_{h,t}^{-k}$	Number of Home Firms
$n_{f,t} = \hat{\varphi}_{f,t}^{-k}$	Number of Foreign Firms
$N_t = n_{h,t} + n_{f,t}$	Total Number of Firms at Home
$C_t = \psi_0^{-\frac{k-(\theta-1)}{\rho(\theta-1)}} \psi_1^{-1} A_t^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}}$	Aggregate Home Consumption
$L_{h,t} = \frac{1}{\theta}\left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta} A_t^{\theta-1} \hat{\varphi}_{h,t}^{-k(\theta-1)}$	Employment at Home Firms
$L_{f,t} = \frac{1}{\theta}\left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta} A_t^{\theta-1} \hat{\varphi}_{f,t}^{-k(\theta-1)}$	Employment at Foreign Firms
$L_t = L_{h,t} + L_{f,t}$	Aggregate Home Employment
$RER_t = \left(\frac{f_f}{f_h}\right)\left(\frac{A_t}{A_t^*}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}}$	Real Exchange Rate
$S_t = \frac{f_f}{f_h}\frac{M_t}{M_t^*}$	Nominal Exchange Rate
$P_t = \psi_0^{\frac{k-(\theta-1)}{\rho(\theta-1)}} \psi_1 \left(\frac{1}{A_t}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}} M_t$	Price Level

Coefficients

$$\psi_0 = \left\{ (\theta f_h)^{\rho} \left(\frac{\theta-1}{\theta}\right) \kappa \left[1 + \left(\frac{f_h}{f_f}\right)^{\frac{\kappa-(\theta-1)}{(\theta-1)}}\right]^{\frac{\rho(\theta-1)-1}{(\theta-1)}} \left(\frac{k}{k-(\theta-1)}\right)^{\frac{\rho(\theta-1)-1}{(\theta-1)}} \right\}^{\frac{\theta-1}{k\rho(\theta-1)+(\theta-1)-k}}$$

$$\psi_1 = \left(\frac{\kappa\theta}{\theta-1}\right)^{\frac{1}{\rho}} \left[1 + \left(\frac{f_f}{f_h}\right)^{\frac{1}{(\theta-1)}}\right]^{-\frac{1}{\rho(\theta-1)}} \left(\frac{k}{k-(\theta-1)}\right)^{-\frac{1}{\rho(\theta-1)}}$$

Aggregate labor supply is found by aggregating over the individual labor supplies:

$$\begin{aligned} L_t &= L_{h,t} + L_{f,t} = \int_0^{n_{h,t}} l_{h,t}(i)di + \int_0^{n_{f,t}} l_{f,t}(i)di = \\ &= \frac{1}{\theta}\left(\frac{\theta-1}{\theta}\frac{1}{\kappa}\right)^{\theta} C_t^{1-\rho\theta} A_t^{\theta-1} \left(\hat{\varphi}_{h,t}^{-k(\theta-1)} + \hat{\varphi}_{f,t}^{-k(\theta-1)}\right). \end{aligned} \tag{36}$$

This expression isolates the three effects working on labor input. Since consumption and leisure are substitutes, the wage increases in aggregate consumption. Unless $\rho < \frac{1}{\theta}$, which would imply that households are close to being risk-neutral, increases in aggregate consumption, driven by productivity increases, reduce labor. A countervailing effect is coming from labor demand,

whereby productivity shocks directly raise employment. The third element is the entry effect identified earlier. Productivity improvements lower the thresholds for both Home and Foreign firms, entry occurs, and labor demand rises.

The consumption effect is generally not strong enough to overturn the direct productivity effect outside of sticky price models, hence the overall effect of productivity shocks on employment is positive. But this is reinforced through the entry mechanism, which implies that in our Melitz-type framework, labor input is likely to be more volatile than in standard models. The reduced-form expression for L_t is straightforward to compute, but lengthy. We thus only report the elasticity of L_t with respect to aggregate productivity: $(\theta - 1) + (1 - \rho)\,\theta \frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}$. It is composed of the direct effect from productivity, $(\theta - 1)$; the second terms amalgamate the indirect effects from consumption-leisure substitutability and entry. In the benchmark case of log-utility, $\rho = 1$ and the indirect effects cancel each other out. When agents are less risk-averse, $0 < \rho < 1$, then the indirect effects amplify labor movements, and have a dampening effect otherwise. We will discuss this insight in more detail below.

The remaining reduced-form solutions are now easy to compute. We forgo discussion of these as they simply reiterate the main themes. The expressions are listed in Table 1. Finally, the model is closed by specifying monetary policy. We assume the money supply evolves according to a simple monetary base rule subject to i.i.d. injections,

$$M_{t+1} = M_t + \varepsilon_{Mt}, \tag{37}$$

where $log\varepsilon_{Mt} \sim N(0, \sigma^2_{\varepsilon_M})$. Seigniorage revenue is rebated to the household: $T_t = M_{t+1} - M_t = \varepsilon_{Mt}$. This completes the specification of the model.

3. DISCUSSION

The logic behind the model emerges most clearly by considering the effects of a productivity shock. We first note that the model contains no endogenous propagation mechanism. Any persistent effects thus stem entirely from serial correlation in the exogenous disturbances. In other words, there are no intertemporal tradeoffs to consider. However, this allows us to cleanly isolate the entry mechanism at play, which is something that is not easily discernible in richer environments built around the Melitz-framework (e.g., Ghironi and Melitz 2005).

Model Mechanics

Suppose aggregate productivity A_t unexpectedly increases by 1 percent. Because overall productivity, composed of aggregate and firm-specific

productivity, rises, firms can expect to generate higher revenue out of which the fixed cost of entry can be more easily financed. The idiosyncratic productivity threshold thus falls for both Home and Foreign-owned firms and entry occurs. If $f_h < f_f$, relatively more Home firms enter, but the overall number of firms in the economy, N_t, increases. The elasticity of the number of firms with respect to productivity can be found by combining equations (26) and (27). This yields an elasticity coefficient of $\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k} > 0$. As it turns out, this is a key coefficient for the behavior of the model. We will analyze its determinants in more detail in the next section.

The flip side of more firms operating in the economy is that it has adverse effects on several productivity measures. Since there are now more lower productivity firms after the positive aggregate productivity shock, average idiosyncratic productivity for home and foreign firms, $\bar{\varphi}_{j,t}$, $j = h, f$, and for the overall economy, $\bar{\varphi}_t$, falls. Vice versa, a decline in aggregate productivity raises average productivity since firm entry declines relative to its steady state. The model thus captures a cleansing effect of recessions and the observed increase in average firm productivity over the course of a downturn. In a similar vein, this also illustrates how measured total factor productivity can be a misleading indicator for actual firm productivity due to the composition effect caused by entry and exit.

The effect on other real quantities is quickly established. The solution for consumption comes directly from the cash-in-advance constraint. Its responsiveness to productivity is again given by the previous coefficient. An increase in aggregate productivity lowers the aggregate price level in equation (31) with the same elasticity coefficient and raises the real exchange rate. As we pointed out before, there is no effect on the nominal exchange rate since the real exchange rate freely adjusts to equilibrate the balance of payment flows generated by the increased FDI from the low to the high productivity country. More Foreign firms enter the domestic market and produce output, which increases $n_{f,t}$. However, the domestic price level falls due to the supply effect, which lowers the nominal value in domestic currency terms of the Foreign-operated firms. As the expressions for the nominal exchange rate show, see equations (30) and (33), these two effects exactly offset each other. We also want to point out that the model preserves monetary neutrality. Money supply shocks only affect the nominal exchange rate, see equation (33).

Entry and Exchange Rate Volatility

We now use the analytical solutions derived above to study the relationship between the nominal exchange rate, the real exchange rate, and the underlying fundamentals. The first issue we discuss is the relationship between the exchange rates and the fundamental shocks, namely the money supply and productivity processes. The background to this discussion is the so-called

exchange rate disconnect puzzle, which stipulates that, empirically, exchange rates appear to behave independently of underlying economic fundamentals—that they are virtually autonomous processes best captured by a unit root model (see Meese and Rogoff 1983). A corollary of this puzzle is that the behavior of real quantities is well captured by underlying shocks, whereas exchange rates are not.

We first note that the nominal and real exchange rates are driven by different shock processes, that is, the dichotomy in this framework between the effects of real and nominal shocks is preserved. Movements in the real exchange rate are explained by movements in relative productivity levels, see equation (32), with an elasticity coefficient of $\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}$. The properties of the underlying driving processes thus carry over to the exchange rates. High persistence in the latter would therefore have to be generated by a high degree of persistence in productivities. One problematic issue is that the underlying shock processes are generally not observable. Consequently, the literature thus often uses the alternative metric of relative consumption. As the expression $\frac{C_t}{C_t^*} = \left(\frac{A_t}{A_t^*}\right)^{\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}}$ shows, this is the same as for the real exchange rate up to a scale factor. Real exchange rates thus move one-to-one with relative consumption ratios. In other words, there is no exchange rate disconnect puzzle in this framework. As in the standard literature with trade in goods, movements in relative consumption are closely tied to the real exchange rate. However, we want to point out again that the only cross-country linkage here is via the capital account in terms of repatriated profits. What proxies for the international risk-sharing condition is thus the balance of payments condition.

We now turn to the other dominant issue in the international macro literature, namely the exchange rate volatility puzzle. There are two aspects to this: one, the relative volatilities of nominal and real exchange rates, and two, the relative volatilities of exchange rates and the underlying shocks. We find it convenient to express the moments in terms of natural logarithms:

$$\begin{aligned} s_t &= const. + m_t - m_t^*, \\ rer_t &= const. + \frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}\left(a_t - a_t^*\right). \end{aligned}$$

Assuming independence amongst the exogenous shock processes, we thus find that the volatility of the exchange rates is given by

$$\begin{aligned} \sigma_s^2 &= \sigma_m^2 + \sigma_{m^*}^2, \\ \sigma_{rer}^2 &= \left[\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}\right]^2 \left(\sigma_a^2 + \sigma_{a^*}^2\right). \end{aligned}$$

As we already pointed out in the discussion above, nominal and real exchange rate movements move independently from each other. It follows that the relative volatilities of the exchange rates are essentially arbitrary in this framework and that the model imposes no restrictions on their co-movement. This is the

outcome of the two arguably extreme assumptions: the lack of international trade in goods and assets (besides profit flows) and the identical nature of both countries. Nevertheless, we regard this result as an interesting benchmark for future literature.

What the model is not silent about, however, is the second aspect of the exchange rate volatility puzzle, namely the degree of amplification of fundamental shocks inherent in the entry mechanism. The key for this is the coefficient:

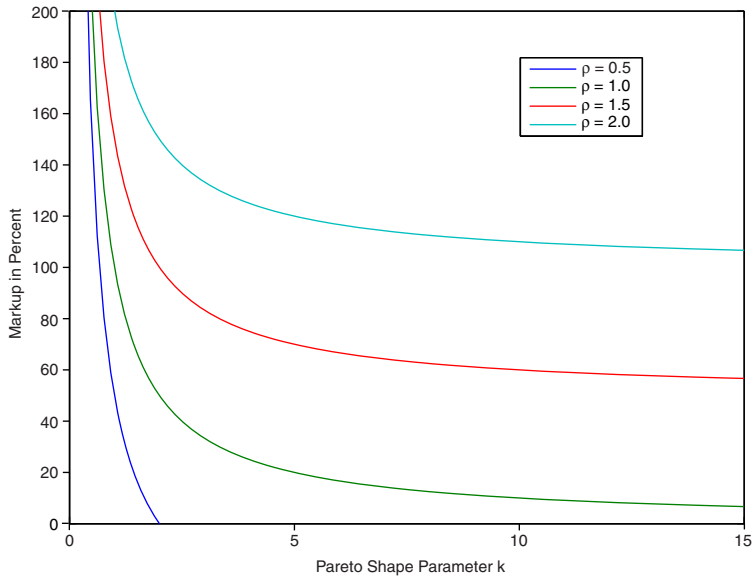
$$\frac{k(\theta-1)}{k\rho(\theta-1)+(\theta-1)-k}\begin{cases} >1 & \text{Amplification} \\ =1 & \text{Equiproportional} \\ <1 & \text{Dampening} \end{cases}.$$

This translates into the following restriction on the parameters: Shocks are amplified (dampened) through the entry mechanism if $\rho < (>)\frac{\theta}{\theta-1} - \frac{1}{k}$. We thus expect productivity shocks to be amplified (i) when θ is low (and the markup high), (ii) when k is large, and (iii) when the degree of risk aversion is low.

We can assess quantitatively whether these conditions are reasonable. Estimates for the shape parameter k of the Pareto distribution and the substitution elasticity θ run the gamut in the literature. Estimates of the dispersion of firm size from Helpman, Melitz, and Yeaple (2004) suggest a value of $k = 11$. Furthermore, di Giovanni, Levchenko, and Rancière (2011) suggest that in any Melitz-model k should be roughly equal to θ. $\theta = 11$ implies a markup of 10 percent, which is an often used value in the macroeconomics literature. On the other hand, Broda and Weinstein (2006) and Feenstra, Obstfeld, and Russ (2011) find values for θ between 2 and 3, which would imply markups between 50 percent and 100 percent.

Figure 1 depicts iso-curves for the equation $\frac{\theta}{\theta-1} = \rho + \frac{1}{k}$, at which there is neither an amplification nor a dampening effect. We report curves for four values of the coefficient of relative risk aversion, $\rho = 0.5,1.0,1.5,2.0$. Areas above and to the right of each curve imply an amplification effect, while below and to the left indicate a dampening effect. It is obvious that an amplification effect generally requires a low degree of risk aversion. This stems from the fact that, with low risk aversion, households willingly substitute into and out of leisure, which implies high labor volatility as we discussed above. At even moderate degrees of risk aversion, for instance, $\rho = 2$, an amplification effect can be ruled out except for implausibly high markups above 100 percent. In a baseline case with log-utility, $\rho = 1$, a value of the shape parameter of $k = 11$, implies a markup of at least 9.1 percent, or $\theta < 12$, for amplification of productivity shocks on real variables; whereas for the alternative case of $\theta = 3$ (and a markup of 50 percent), a value of $k > 2$ would be required. None of these baseline cases appear implausible. In fact, a markup of 10 percent and log-utility is quite standard in the macro literature. However, they are predicated on a narrow range for the risk-aversion parameter. Any

Figure 1 Amplification Iso-Curves



Notes: The graph depicts the iso-curves $\frac{\theta}{\theta-1} = \rho + \frac{1}{k}$ for various values of the coefficient of relative risk aversion ρ. Areas above and to the right of each curve imply an amplification effect, while below and to the left indicate a dampening effect.

amplification that occurs can be sizeable, however. For instance, when $\rho = 1$, $k = 11$, and $\theta = 4$ (implying a markup of 33 percent), the amplification effect is 32 percent. Given the stylized nature of the model, this appears to us as quite large.[8]

Testable Implications

Given the nature of the quantitative exercise above, any potential empirical statements would have to be heavily qualified. However, the analysis yields several interesting testable implications regarding when amplification effects are most likely to be important. First, Figure 1 shows that the lower is the Pareto shape parameter k, the greater is the range of elasticities for which

[8] We should point out a further caveat to our analysis. The various exchange rate puzzles are typically discussed for high frequency data of a quarter or less. In our framework, the time period is arguably of a much lower frequency since the FDI process of physically locating production abroad takes place on a longer time scale.

amplification effects arise. Thus, we would expect a generally positive causal relationship between multinational firm activity and the relative volatility of the exchange rates for countries with a large degree of multinational activity in industries with a higher dispersion in firm size (that is, a low k) and thus higher industry concentration.

Second, it is apparent that for countries with FDI in manufacturing sectors focused on the production of products with high markups (that is, highly differentiated goods with a low elasticity of substitution), amplification effects are much more likely, even with higher measures of market concentration indicative of low k. Finally, regardless of the size of these parameters, countries and industries with higher fixed costs for multinationals relative to domestic firms (high $\frac{f_f}{f_h}$) will exhibit greater amplification effects. Higher fixed costs may arise due to difficulties connected with obtaining crucial information about the host market, communicating and coordinating with headquarters, or surmounting technological hang-ups. Thus, all else equal, we would expect excess volatility stemming from multinational firm activity to be decreasing in the quality of a country's infrastructure and institutions, and increasing in the level of technological sophistication of its main manufacturing sectors in which FDI plays a key role.[9]

In short, the most promising avenue for the Melitz-type framework we developed to make a contribution to the international trade and macro literature is through an amplification effect of shocks and a variable entry and exit mechanism. The quantitative importance of this mechanism rests on a narrow (though commonly used) set of parameter values within the boundary of what is likely empirically founded. Our quantitative analysis points to three testable implications for researchers seeking to investigate the causal link between multinational activity and excess volatility.

4. CONCLUSION

We build a simple model of market entry with heterogeneous firms and multinational production. We are able to characterize the solutions for all variables analytically, which allows us to identify the key mechanism in the model without having to resort to numerical methods. Fluctuations in the net profits repatriated by multinational firms can generate real and nominal exchange rate volatility. Variability in repatriated profits, since it is entirely dependent upon consumption in our Melitz-type framework with homothetic preferences and constant markups, does not generate a disconnect—variability in the real

[9] We note that when the Pareto shape parameter k is less than 2.5, as is the case in estimates for all industries by di Giovanni, Levchenko, and Rancière (2011), the degree of risk aversion is not a prime determinant of whether amplification effects arise due to multinational behavior. Thus, the degree of risk aversion should be of second-order importance in an empirical analysis of the causal effects of FDI on excess volatility.

exchange rate is driven by exactly the same factors and to the same degree as relative consumption. However, there is a potential for disconnect between the real and the nominal exchange rate: the first is driven by productivity shocks and the second by monetary shocks.

In addition, we derive conditions under which the volatility of the real exchange rate can deviate from the volatility of underlying productivity shocks, dampened or amplified by the entry and exit and profit remittances of multinational firms. A reasonable range of parameters can produce either effect. Amplification, that is, excess volatility, emerges under the most commonly used set of parameters, which is remarkable in that it occurs even though prices are fully flexible and markups are constant. In particular, we find that excess volatility in our flexible-price framework is most likely when the distribution of firm size is more fat-tailed, when industries in which FDI is important are highly differentiated with high levels of technical sophistication that generate large coordination costs specific to multinationals, and in countries with low levels of infrastructure and institutional development. In this way, we link a macroeconomic puzzle to the microeconomics of industry structure using the tools from the New Trade Theory.

REFERENCES

Bacchetta, Philippe, and Eric van Wincoop. 2000. "Does Exchange-Rate Stability Increase Trade and Welfare?" *American Economic Review* 90 (December): 1,093–109.

Backus, David K., Patrick J. Kehoe, and Finn E. Kydland. 1992. "International Real Business Cycles." *Journal of Political Economy* 100 (August): 745–75.

Baxter, Marianne, and Alan C. Stockman. 1989. "Business Cycles and the Exchange-Rate Regime: Some International Evidence." *Journal of Monetary Economics* 23 (May): 377–400.

Bernard, Andrew B., J. Bradford Jensen, and Peter Schott. 2009. "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. Chicago: University of Chicago Press, 513–52.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *The Quarterly Journal of Economics* 121 (May): 541–85.

Cavallari, Lilia. 2007. "A Macroeconomic Model of Entry with Exporters and Multinationals." *The B.E. Journal of Macroeconomics* 7 (1).

Clarida, Richard, and Jordi Galí. 1994. "Sources of Real Exchange Rate Fluctuations: How Important are Nominal Shocks?" *Carnegie-Rochester Conference Series on Public Policy* 41 (December): 1–56.

Di Giovanni, Julian, Andrei A. Levchenko, and Roman Rancière. 2011. "Power Laws in Firm Size and Openness to Trade: Measurement and Implications." *Journal of International Economics* 85 (September): 42–52.

Feenstra, Robert C., Maurice Obstfeld, and Katheryn N. Russ. 2011. "In Search of the Armington Elasticity." Working Paper, University of California, Davis.

Fillat, José L., and Stefania Garetto. 2010. "Risk, Returns, and Multinational Production." Working Paper, Boston University.

Fitzgerald, Doireann. 2008. "Can Trade Costs Explain Why Exchange Rate Volatility does not Feed Into Consumer Prices?" *Journal of Monetary Economics* 55 (April): 606–28.

Flood, Robert P., and Andrew K. Rose. 1995. "Fixing Exchange Rates: A Virtual Quest for Fundamentals." *Journal of Monetary Economics* 36 (August): 3–37.

Ghironi, Fabio, and Marc J. Melitz. 2005. "International Trade and Macroeconomic Dynamics with Heterogeneous Firms." *The Quarterly Journal of Economics* 120 (August): 865–915.

Gourinchas, Pierre-Olivier, and Hélène Rey. 2007. "International Financial Adjustment." *Journal of Political Economy* 115 (4): 665–703.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export Versus FDI with Heterogeneous Firms." *American Economic Review* 94 (March): 300–16.

Kosteletou, Nikolina, and Panagiotis Liargovas. 2000. "Foreign Direct Investment and Real Exchange Rate Interlinkages." *Open Economies Review* 11 (April): 135–48.

Lubik, Thomas A., and Frank Schorfheide. 2005. "A Bayesian Look at New Open Economy Macroeconomics." In *NBER Macroeconomics Annual 2005*, edited by Mark Gertler and Kenneth Rogoff. Cambridge, Mass.: MIT Press, 313–66.

Meese, Richard A., and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14 (February): 3–24.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (November): 1,695–725.

Phillips, Shauna, and Fredoun Z. Ahmadi-Esfahani. 2008. "Exchange Rates and Foreign Direct Investment: Theoretical Models and Empirical Evidence." *Australian Journal of Agricultural and Resource Economics* 52 (December): 505–25.

Ramondo, Natalia, Veronica Rappoport, and Kim J. Ruhl. 2010. "The Proximity-Concentration Tradeoff under Uncertainty." Working Paper, Arizona State University.

Rogoff, Kenneth S. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34 (June): 647–68.

Russ, Katheryn N. 2007. "The Endogeneity of the Exchange Rate as a Determinant of FDI: A Model of Entry and Multinational Firms." *Journal of International Economics* 71 (April): 344–72.

Russ, Katheryn N. 2011. "Exchange Rate Volatility and First-Time Entry by Multinational Firms." *Review of World Economics* 148 (2): 269–95.

Shrikhande, Milind M. 2002. "A General Equilibrium Analysis of Foreign Direct Investment and the Real Exchange Rate." *International Journal of Finance and Economics* 7 (October): 309–25.