

Collateral-Motivated Financial Innovation*

Ji Shen
London School of Economics

Hongjun Yan
Yale School of Management

Jinfan Zhang
Yale School of Management

March 27, 2012

*We thank John Geanakoplos, Andrew Metrick, Adriano Rampini and seminar participants at MIT and Yale, as well as Econometric Society Meetings, University of Minnesota Finance Conference for helpful comments. Please direct all correspondence to Hongjun Yan, Email: hongjun.yan@yale.edu. The latest version of the paper is available at <http://faculty.som.yale.edu/hongjunyan/>.

Collateral-Motivated Financial Innovation

Abstract

We propose a collateral view of financial innovation: Many innovations are motivated by alleviating collateral/margin constraints for trading (speculation or hedging). We analyze a model of investors with heterogeneous beliefs. The trading need motivates investors to introduce derivatives, which are endogenously determined in equilibrium. In the presence of a collateral friction in cross-netting, the “optimal” security is the one that isolates the variable with disagreement. It is optimal in the sense that alternative derivatives cannot generate any trading. With an arbitrarily small trading cost, the optimal security is “unfunded”, i.e., has a zero initial value. The endogenous difference in collateral requirements leads to a basis, i.e., the spread between the prices of an underlying asset and its replicating portfolio. This basis reflects the shadow value of collateral, leading to a number of time-series and cross-sectional implications. More broadly, our analysis highlights the common theme behind a variety of financial innovations: the inventions of securities (e.g., futures, swaps); legal practice (e.g., the superseniority of repos and derivatives); legal entities (e.g., special purpose vehicles); as well as the efforts in improving the margin procedure.

JEL Classification Numbers: G11, G23.

Keywords: Financial Innovation, Collateral, Margin, Leverage, Asset Pricing, Derivative, Basis.

1 Introduction

The last half a century has witnessed a tremendous amount of financial innovations. What are the motives behind them? Existing theories emphasize the role of risk sharing (e.g., Allen and Gale (1994)), transaction costs and regulatory constraints (Benston and Smith (1976), Miller (1986)) and information asymmetry (Gorton and Pennacchi (1990), DeMarzo and Duffie (1999)).

This paper proposes an alternative view: many successful financial innovations are partly motivated by mitigating collateral (or margin) constraints for trading. Suppose, for example, two traders have different expectations on the future value of a security, say, a corporate bond. If their disagreement is about the company's default probability, rather than the future movements of riskless interest rates, then it is natural that the traders prefer to take positions in credit default swaps (CDS), rather than the corporate bond. This is because, by isolating the default probability, the variable that traders are interested in betting on, CDS requires least collateral and is efficient in facilitating their speculation.¹ This collateral motivation is not limited to speculative trading: Suppose, for instance, a risk manager of a corporation needs to hedge a certain exposure, and can trade two financial instruments with the same hedging quality. To the extent that raising capital is costly, the risk manager clearly has a preference for the instrument with a lower collateral requirement.

Motivated by the above intuition, we analyze an equilibrium model of investors with heterogeneous beliefs about a portion of a cash flow from an asset. The disagreement motivates investors to trade this asset, and possibly to introduce new derivatives to facilitate their trading. Casual intuition suggests that investors would introduce derivatives that are linked to the disagreement. However, it is far less clear about the impact of this innovation on other markets. Would investors try to complete the markets? Which markets would thrive, and which would disappear? What is the notion of "optimal" innovation in this context?

To understand these issues, let's first consider a benchmark case without collateral frictions.

¹A vivid example is documented in Michael Lewis's book *Big Short*. During 2004-2006, a number of investors were convinced that the subprime mortgage market would soon collapse, and wanted to bet on it. However, they found existing instruments (e.g., the stocks of home building companies) can only offer an "indirect" bet. The book tells a detailed story of how those investors push banks to create the market of CDS contracts on subprime mortgage bonds, which provides a more "direct" bet on the subprime mortgage market.

In this case, if an investor defaults on his promise (e.g., debt or a short position in an Arrow security), his counterparty can seize the collateral the investor has posted for the trade and the defaulting investor faces no further penalty.² The benchmark case features a frictionless collateralization procedure, in which an investor can use (any part of) his overall portfolio as collateral. For convenience, we refer to it as *portfolio margin*. It is easy to see that the collateral constraint under portfolio margin is equivalent to a nonnegative wealth constraint. Moreover, if investors introduce financial assets to complete markets, the resulting equilibrium is Pareto optimal. This benchmark case highlights the benefit of market completeness but does not have sharp predictions on financial innovation.

Our main analysis is focused on a collateral friction. In particular, we note that portfolio margin is often impractical. For example, if an investor holds a portfolio in which individual asset returns offset each other, under portfolio margin, the collateral requirement for the whole portfolio can be much lower than that for one individual asset alone. In practice, however, this cross-asset netting is far from perfect. For instance, if one asset in the portfolio is exchange-traded while the other is over-the-counter or traded at a different exchange, then the investor has to post collateral for both assets separately, even if these two positions largely offset each other.³ Moreover, it may be hard or too costly for a dealer to precisely estimate the correlation among securities to determine the collateral for the whole portfolio. Or a trader may prefer not to reveal his whole portfolio to his dealer by having multiple dealers, which is a common practice among hedge funds. Finally, different parts of the portfolio may be governed by different jurisdictions and regulations may also impose various constraints on collateralization, making cross-netting imperfect.

Our key assumption, motivated by these frictions in collateralization, is that investors in our model have to post collateral for each security in their portfolios separately, which we refer to

²This lack of further penalty assumption is perhaps most suitable for thinking about the case of security trading, where many positions are set up for hedging, speculation or short term financing purposes. When an investor defaults, the top priority for his counterparties is perhaps to get compensated quickly to reestablish those positions with other investors, rather than going through a lengthy (e.g., bankruptcy) procedure to get compensated by liquidating the defaulting investor's other assets.

³One famous example is that Metallgesellschaft AG, a German conglomerate, had a large short position in oil forward and an offsetting long position in oil futures in early 1990s, but eventually ran into liquidity crisis when the collateral requirement became excessive (see, Culp and Miller (1995)).

as *individual security margin*. It is worth noting that the essential point of this assumption is that cross-asset netting is imperfect rather than impossible. Our model's main implications are the following.

First, this collateral friction determines which financial innovation would be adopted in equilibrium. Intuitively, due to the collateral friction, rather than trading the underlying asset, investors prefer to trade a derivative that isolates the portion of the cash flow with disagreement. This is because the cash flow from the underlying asset has two portions but investors are only interested in trading one. To the extent that the "unwanted" portion, the portion without disagreement, increases the collateral requirement for trading the underlying asset, it makes the underlying asset less appealing than the derivative. Consequently, the derivative that completely carves out the unwanted cash flow is "optimal" in the sense that its existence would drive out any other derivative markets: if one introduced any other derivatives, those markets would not generate any trading.

Second, the optimal derivative tends to be "unfunded," i.e., the initial value of the derivative is designed to be zero. The reason is that for a security that facilitates speculation or hedging, its essential role is to transfer wealth across states in the future, rather than across time. Hence, if its price is not zero, one party gets paid initially, but there is a chance for him to pay this amount back in the future. Making the security unfunded avoids this potential "round trip" in wealth transfer. To the extent there is an infinitesimal cost of transferring funds, unfunded security would be strictly preferred. This perhaps explains why many derivatives, such as futures and swaps, are designed to be unfunded.

Third, due to the high collateral requirement, the price of the underlying asset can be lower than the price of its replicating portfolio. This is consistent with the empirical evidence on the so-called corporate bond-CDS basis: the price of a corporate bond is often lower than the price of the portfolio of a CDS and a Treasury bond that replicates the corporate bond's cash flow (e.g., Mitchell and Pulvoni (2011), Garleanu and Pedersen (2011)). More recently, Fleckenstein, Longstaff, and Lustig (2010) find that the prices of Treasury Inflation-Protected Securities (TIPS) are lower than that of their replicating portfolios that consist of inflation

swaps and nominal Treasury bonds. These phenomena are consistent with our model: It takes more collateral to take a long position in the underlying assets (i.e., corporate bonds or TIPS). Even for TIPS, the collateral requirement is around 3% in Fleckenstein, Longstaff, and Lustig (2010). In contrast, to take an equivalent long position through those unfunded derivatives (i.e., inflation swaps and CDS), one just need to post collateral to cover the daily movements in the mark-to-the-market value, and so the collateral requirement is much smaller. This creates a basis, the spread between the price of the underlying asset and its replicating portfolio.

Forth, the above intuition implies that the basis reflects the shadow value of collateral, leading to a number of time-series and cross-sectional implications on the spread. For example, when investors face tighter funding constraints, saving collateral becomes more valuable, leading to a larger basis. This is consistent with the evidence that both the corporate bond-CDS basis and TIPS-inflation-swap basis increased dramatically during the financial crisis in 2007-2008, when investors perhaps were facing tighter funding constraints. Our model also implies that the basis is higher if the unwanted cash flow, the portion of the cash flow that is carved out by the derivative, is more volatile. This is because the unwanted cash flow volatility determines how much collateral can be saved by trading the derivative rather than the underlying. Moreover, when investors' funding liquidity dries up, the basis increases more for assets with more volatile carved-out cash flows. There has been some evidence supporting these implications. For example, Mitchell and Pulvino (2010) find that, during the crisis, the corporate-bond-CDS basis tends to be larger for junk bonds than for investment grade bonds. Moreover, with the financial crisis unfolding, the basis for junk bonds increases more than that for investment grades. These results are consistent with our model predictions, if one takes the interpretation that the carved-out cash flows for junk bonds are more volatile (e.g., perhaps due to liquidity risk). Our model also provides a number of new testable predictions. For example, it implies that the basis should be higher for corporate bonds and TIPS with longer maturities, or when there is more interest rate uncertainty. The basis increases when there is a positive supply shock to the underlying asset (e.g., when a failing institution has to sell a large amount of the underlying asset), or when there is an increase in trading need. The basis for corporate bonds and TIPS with longer maturities should increase more when there is a liquidity shortage or supply shock to the underlying assets.

Finally, and more broadly, this collateral view of financial innovation highlights the common theme behind a variety of financial innovations with strikingly different appearances. For example, many successful derivative contracts are unfunded and allow investors to take on large positions with very little collateral. In addition to the previously mentioned swaps, futures contract is another example. A futures contract allows investors to take an exposure to the fluctuations of the price of a certain asset without the physical process of buying or selling the asset. This is especially important for commodity futures where transactions are costly and time consuming. In other words, similar to the case for swaps, futures contracts save collateral by isolating the variables that investors want to bet on.

Another example of collateral-motivated innovation is the emerging legal practice of the so-called superseniority of derivatives and repos. Although derivatives and repos are not supersenior in a strict statutory sense, it has been a common practice in the U.S.: When a company goes bankrupt, its repo and derivative counterparties can simply seize the collateral posted in the transactions up to the amount the company owes them, instead of going through the lengthy and costly bankruptcy procedure.⁴ In the context of our analysis, this practice can be viewed as carving unwanted cashflows out of derivative and repo transactions: Suppose an investor enters an interest rate swap and his goal is to hedge or speculate on interest rate risk. Without superseniority, even if his counterparty posts a large amount of collateral, the investor is still not well protected since he would have to go through automatic stay when his counterparty defaults. With superseniority, however, the investor can immediately seize the collateral upon default, and so can be better protected even by a smaller amount of collateral. To see how the efficiency is achieved, note that given the investor's purpose, his counterparty's assets, apart from the posted collateral, are unwanted cashflows, and are carved out of the swap transaction by superseniority. Similarly, financial innovation may take the form of new legal entities. For example, special purpose vehicles (SPVs), have become prevalent in recent decades with the rise of securitization. Again, in the context of our analysis, we can view creating an SPV as carving

⁴This exceptional treatment accorded derivatives and repos in bankruptcy is recent. It was formalized by the introduction of "Act to Amend Title 11, United States Code, to Correct Technical Errors, and to Clarify and Make Substantive Changes, with Respect to Securities and Commodities" to the bankruptcy code as of July 27, 1982 (PL 97-222 (HR 4935)). There have been numerous revisions over the years. A recent example is the Financial Netting Improvements Act of 2006 (Pub. L. 109-390).

out unwanted cashflows (i.e., the firm's assets other than those allocated to the SPV).

Note also that the collateral friction in our model arises from the limitations on cross-netting in posting collateral. In practice, it is becoming increasingly possible to have more cross-netting. For example, on December 12 2006, the Securities and Exchange Commission (SEC) approved a rule change which made partial cross-netting available to some investors in the exchange-traded options market.⁵ There have also been efforts from brokers and hedge funds that attempt to get around the regulation-induced margin requirements (see, e.g., Brunnermeier and Pedersen (2009)). One can view these continuing efforts by regulators and market participants in modifying the margin procedure as one form of collateral-motivated financial innovation. Their goal is simply to satisfy the demand from market participants to alleviate their collateral constraints.

2 Literature Review

There is an extensive literature on financial innovation. Recent surveys, such as Allen and Gale (1994), Duffie and Rahi (1995) emphasize the risk sharing role of financial instruments, Tufano (2003) also discusses the roles of regulatory constraints, agency concerns, transaction cost and technology (Ross (1989), Benston and Smith (1976), Merton (1989), White (2000)). Some more recent studies explore the role of rent seeking (Biais, Rochet and Woolley (2010),) and neglected risk (Gennaioli, Shleifer and Vishny (2010)). These studies generally abstract away from collateral constraint, which is the focus of this paper. One exception is Santos and Scheinkman (2001), which analyze a model where exchanges set margin levels to screen traders with different credit qualities. Also related are the studies that analyze the impact of financial innovation in models with heterogeneous beliefs or preferences (Zapatero (1998), Bhamra and Uppal (2009), Simsek (2011), Banerjee and Graveline (2011)). These studies focus on the impact of innovation on volatility and the underlying asset price, while our analysis focuses on the collateral friction and its implications on endogenous financial innovation and asset prices.

The role of collateral has been analyzed in various contexts, such as macro economy (e.g.,

⁵For more details see the *Customer Portfolio Margin User Guide*, available from the website of The Options Clearing Corporation, www.optionsclearing.com.

Kiyotaki and Moore (1997)), corporate debt capacity (e.g., Rampini and Viswanathan (2010)), arbitrageur's portfolio choices (e.g., Liu and Longstaff (2004)), asset prices and welfare (Basak and Croitoru (2000), Gromb and Vayanos (2002)). Our analysis on collateral requirement builds on earlier work of Geanakoplos (1997, 2003), which has been extended to study leverage cycle (Fostel and Geanakoplos (2008), Geanakoplos (2009)), speculative bubble (Simsek (2011)) and debt maturity (He and Xiong (2010)). Our analysis on leverage is also related to the studies of financial products that help constrained investors to take leverage (Frazzini and Pedersen (2011), Jiang and Yan (2012)).

Finally, our paper is closely related to Garleanu and Pedersen (2011), who analyze the impact of collateral on the violation of the law of one price. A major difference is that the endogenous financial market structure and collateral requirements are the focus of our paper, but are exogenously given in their study. While many implications on basis are similar across these two studies, our model also has new predictions, e.g., the impact of supply shocks on basis. There is also a more subtle difference: Suppose there are two portfolios with identical cash flows. While both studies show that the one with a lower collateral requirement has a higher price, our analysis further suggests that it is not a coincidence that derivatives tend to be in the portfolio that demands *less* collateral—that's the point of inventing those derivatives!

The rest of the paper is organized as follows. Section 3 presents the model and its equilibrium. Section 4 analyzes the violation of the law of one price. Some further analysis of the model is presented in Section 5. Section 6 studies the impact of financial innovation on the economy. Section 7 provides some general discussions and Section 8 concludes. All proofs are summarized in the Appendix.

3 A Model of Financial Innovation

We consider a two period economy, $t = 0, 1$, which is populated by a continuum of investors. The total population is normalized to 1. Investors make portfolio decisions at $t = 0$ and consume all their wealth at $t = 1$. All investors are risk neutral and their objective is to maximize their expected consumption at $t = 1$. There is a riskless storage technology with a return of 0. All

investors have the same endowment and the aggregate endowment is e ($e \geq 0$) dollars in cash and β ($\beta \geq 0$) unit of asset A , which is a claim to a random cash flow \tilde{A} at $t = 1$. Investors have different beliefs about the distribution of the cash flow and the disagreement is focused on a portion of it. More precisely, we denote the cash flow as

$$\tilde{A} = \tilde{V} + \tilde{U}, \quad (1)$$

and investors disagree on the distribution of \tilde{V} but share the same belief about the distribution of \tilde{U} . We assume \tilde{V} has a binary distribution. There are two types of investors, optimists o and pessimists p . Investor i , $i \in \{o, p\}$, believes the distribution of \tilde{V} is

$$\tilde{V} = \begin{cases} V_u & \text{with a probability } h_i, \\ V_d & \text{otherwise,} \end{cases} \quad (2)$$

with $V_u > V_d$ and $h_o > h_p$. We use α_o and α_p to denote the population sizes of optimistic and pessimistic investors, respectively and $\alpha_o + \alpha_p = 1$.

Without loss of generality, we assume \tilde{U} has a mean of zero. In addition, we have the following simplifying assumptions: First, \tilde{U} has a bounded support on $[-\Delta, \Delta]$, with $\Delta > 0$ and $V_u - \Delta > V_d + \Delta$. Second, $V_d - \Delta \geq 0$, i.e., \tilde{A} is nonnegative. Third, \tilde{U} is independent of \tilde{V} . We use $F(\cdot)$ to denote the cumulative distribution function of \tilde{U} , and assume that $F(\cdot)$ is differentiable. It is straightforward to generalize these assumptions and the analysis remains similar but becomes more tedious.

Investors agree to disagree, and hence have the incentive to trade among themselves. Naturally, optimistic investors want to buy asset A , and pessimistic investors want to sell. The focus of our analysis is, given the trading motive, how financial innovation would facilitate their trading and affect asset prices and investors' welfare.

3.1 Speculation v.s. Hedging

In the above discussion, investors' trading is motivated by speculation. However, one can easily reinterpret the model so that the trading is motivated by hedging. For example, one interpretation is that all investors have rational expectations but have different preferences. The utility

function of investor i , $i \in \{o, p\}$, is given by

$$u_i(c) = \begin{cases} h_i c & \text{if } \tilde{V} = V_u, \\ (1 - h_i)c & \text{if } \tilde{V} = V_d. \end{cases} \quad (3)$$

One can interpret this specification as investor o having some hedging need at the “up state” $\tilde{V} = V_u$. For example, investor o may incur some extra cost (e.g., the cost of financial distress) at the up state, and so has an incentive to hedge this risk. That is, he has a relatively high marginal utility, h_o , at the up state. Likewise, investor p has a relatively high marginal utility at the “down state.” This modeling device is similar to that in liquidity provision models in which some investors prefer “early” consumptions while others prefer “late” ones: Specification (3) implies that some investors prefer consumption at the up state while others prefer consumption at the down state. Note that, in our model, the speculation and hedging interpretations are mathematically equivalent. In our later discussions, we will mostly adopt the speculation interpretation, and it is straightforward to restate the results under the hedging interpretation.

3.2 Default

Following Geanakoplos (1997, 2003), we assume that, upon default, the debt holder (or derivative counterparty) can seize the collateral posted in the trade, but the defaulting investor faces no further penalty. This assumption can be broadly interpreted as limited enforcement.⁶ Essentially, our assumption implies that when an investor defaults, his counterparty can only seize the collateral posted for his trade, and finds it too costly to get further compensation by penalizing the defaulting investor (e.g., seizing other assets). Therefore, our analysis is perhaps best suitable for security trading, where, in the event of default, the top priority for creditors is to get compensated quickly. In the Lehman Bankruptcy case, for example, 80% of Lehman’s derivatives counterparties terminated their contracts within weeks of bankruptcy. That is, if the derivative position is in-the-money for a counterparty, this company can immediately seize Lehman’s collateral in its margin account for that trade. If the collateral value is less than the amount owed to the company, however, it would be very costly for the company to seize other assets, because it has to go through the lengthy bankruptcy procedure as an unsecured

⁶See Kehoe and Levine (1993) for an early contribution. This idea has lately been applied to asset pricing, see, e.g., Alvarez and Jermann (2000), Chien and Lustig (2009).

debt holder. For example, the final settlement plan for Lehman bankruptcy was approved more than three years later in November 2011 and the senior bondholders only get 21.1 cents on the dollar. Even more seriously, the cost is not only the time value, because many of the over 906,000 derivatives transactions are likely to be for hedging, speculation, or short-term financing purposes. For Lehman's counterparties, the failure to get compensated quickly to reestablish those positions with other counterparties is likely to be much more costly.⁷

This lack of penalty upon default implies that investors need to post collateral to back up their promises (i.e., short positions in future cash flows). The focus of our analysis is how the collateralization process, more precisely the friction in it, determines the financial innovation in equilibrium. In the following, before introducing the collateral friction, we first consider a frictionless benchmark.

3.3 Benchmark Case

Let's first consider the benchmark case with a perfect collateralization procedure. Specifically, an investor can use any of his asset as collateral. For convenience, we refer to this frictionless collateralization procedure as "portfolio margin" since an investor only needs to post collateral for his overall portfolio. It is straightforward to define the *collateral equilibrium with portfolio margin* as the prices of asset A and all derivatives introduced, each investor's positions in the riskless technology, asset A and derivatives, such that all investors' positions satisfy the portfolio margin constraint; all investors maximize their expected utility; and all financial markets clear: the aggregate holding in asset A is β and the aggregate holding in each derivative is zero.

Collateral constraints generally put restrictions on investors' trading. For example, investors cannot borrow without collateral. How does this constraint affect equilibrium prices? With this perfect collateralization procedure, it is easy to see that the collateral constraint is equivalent to the constraint that an investor's wealth has to be nonnegative for all possible states at $t = 1$. That is, this collateral constraint only rules out "empty" promises. The collateral equilibrium is identical to the equilibrium without collateral constraints but investors face a non-negative

⁷All the numbers for Lehman bankruptcy are based on Bala Dharan's speech at NYU Stern Five-Star Conference on Research in Finance on December 2, 2011.

wealth constraint at $t = 1$.

Without collateral frictions, investors can simply invent a complete set of Arrow securities in this economy, so they can achieve the same equilibrium allocation and prices as in the traditional complete market equilibrium, which is Pareto optimal. However, this benchmark case does not have a sharp prediction on which market will be developed in equilibrium. Casual intuition suggests that investors would introduce derivatives that are linked to the disagreement. However, it is far less clear what would happen to other derivative markets. Would investors try to complete the markets? Which markets would thrive, and which would disappear? What is the notion of “optimal” innovation in this context? To shed light on these issues, we need to take seriously the frictions in the collateralization procedure.

3.4 Economy with Individual Security Margin

The key collateral friction analyzed in this paper is that an investor has to post collateral for each position in his portfolio separately, which we refer to as “individual security margin.” Note that the collateral requirement under “portfolio margin” can be much smaller than that under “individual security margin.” For example, if an investor holds a portfolio in which individual asset returns offset each other, under portfolio margin, the collateral requirements for the whole portfolio can be much lower than that for one individual asset alone.

In practice, however, this cross-asset netting is far from perfect. For example, if one asset in the portfolio is exchange-traded while the other is over-the-counter or traded at a different exchange, then the investor has to post collateral for both assets separately, even if these two positions largely offset each other. One famous example is that Metallgesellschaft AG, a German conglomerate, had a large short forward position in oil and an offsetting long position in oil futures in early 1990s, but eventually ran into liquidity crisis when the collateral requirement became excessive (see, Culp and Miller (1995)). Moreover, it may be hard or too costly for a dealer to precisely estimate the correlation among securities to determine the collateral for the whole portfolio. Or a trader may prefer not to reveal his whole portfolio to his dealer by having multiple dealers, which is a common practice among hedge funds. Finally, different parts of

the portfolio may be governed by different jurisdictions and regulations may also put various constraints on collateralization. For example, the Board of Governors of the Federal Reserve System has a number of regulations on the initial margin requirements (Regulations T, U, X) for various institutions.

Our individual security margin assumption captures these frictions by ruling out cross-netting, which manifests itself on the following occasions: 1) When an investor takes a long position in an asset, he can use the asset itself as collateral to borrow to finance the purchase. However, an investor cannot use one risky asset as collateral to finance the purchase of another risky asset, or to issue state contingent debts to finance the purchase of a risky asset. 2) When an investor shorts an asset, he needs to put the proceedings as well as some of his own capital in cash into the margin account. This cash collateral requirement means that investors cannot use an risky asset as collateral to short another risky asset.

It is easy to see that cross-netting is necessary if investors are allowed to use one risky asset as collateral to long or short another risky asset, or to issue state contingent debts to finance the purchase of a risky asset. These nonstandard procedures are therefore more costly and our assumption rules them out for simplicity. The essence of the assumption is the cross-netting is imperfect rather than impossible.

These assumptions closely reflect the practice in reality. For example, to purchase securities on margin is to use those securities themselves as collateral and margin loans are generally not state contingent. The securities are placed in the margin account in “street name”, i.e., the broker-dealers are the legal owners and can lend those securities for short sale by other customers and can liquidate those positions when investors fail to main certain margin requirements (see, e.g., Fortune (2000)). On the short side, as noted by Geczy, Musto and Reed (2002), the collateral for equity loans is almost always cash, and the standard collateral for U.S. equities is 102% of the shares’ value. In summary, our model is perhaps more suitable for analyzing security trading, rather than corporations using their assets as collateral to borrow to finance their investments.

Finally, one might think that long and short positions are treated asymmetrically in our

model: one can only use cash as collateral for short positions but there is no such constraint for the collateral for long positions. This may appear problematic for derivative contracts, where the distinction between long and short positions is immaterial. For example, a long position in a derivative that is a claim to a cash flow $-\tilde{A}$ is the same as a short position in a derivative that is a claim to \tilde{A} . To see this, note that the derivative $-A$ has a negative price and hence an investor is “paid” when taking a long position. Moreover, using this derivative as collateral, an investor can only “borrow” a negative amount (i.e., the investor has to put up some of his own cash). In other words, a levered long position in this derivative $-A$ is the same as a short position in derivative A —relabeling long and short positions has no real impact.

3.5 Equilibrium with Individual Security Margin

We define the probability space spanned by \tilde{V} and \tilde{U} as $\mathcal{H} \equiv (\{V_d, V_u\} \times [-\Delta, \Delta], \mathcal{F}, \mathcal{P}_o, \mathcal{P}_p)$, where $\{V_d, V_u\} \times [-\Delta, \Delta]$ is the sample space, \mathcal{F} is the sigma-algebra generated by $\{V_d, V_u\} \times [-\Delta, \Delta]$, \mathcal{P}_o and \mathcal{P}_p are the probability measures for the optimists and pessimists, respectively. A financial security in this economy is a claim to a cash flow that can be described by a random variable \tilde{K} in \mathcal{H} . We simply refer to this security as “asset K ”.

We now describe formally the investment opportunity sets faced by investors, and define the equilibrium in the presence of derivative K . If an investor takes a long position in an asset, he can use this asset as collateral to borrow to finance part of the purchase. We use (L, C) to denote the borrowing contract, where L is the borrowing amount and the collateral of this borrowing is one unit of asset C ($C = A$ or K). Denote the notional interest rate of this borrowing as $r(L, C)$. Therefore, at time $t = 1$, the lender receives $\min(L(1 + r(L, C)), \tilde{C})$, where \tilde{C} is the value of asset C at $t = 1$. That is, the lender receives $L(1 + r(L, C))$ when there is no default, and he seizes the collateral asset C if the borrower defaults (i.e., when $\tilde{C} < L(1 + r(L, C))$).

We use (C, θ, L) , with $\theta \geq 0$, to denote a long position of θ units of asset C , which is financed by the borrowing contract (L, C) . That is, the investor buys θ units of asset C and use the position itself as collateral to borrow θL to finance the purchase. The payoff of this position

at $t = 1$ is

$$W^+(C, \theta, L) = \max(\theta(\tilde{C} - L(1 + r(L, C))), 0).$$

That is, if the asset value is enough to pay back the debt, the investor gets the residual value. Otherwise, the investor defaults and gets zero. Note that, the investor borrows θL to purchase an asset that is worth θP_C , and so this position needs $\theta(P_C - L)$ capital from the investor, where P_C is the price of asset C at $t = 0$. Similarly, if an investor lends θL and take θ units of asset C from the borrower as collateral, his payoff at time $t = 1$ is

$$X(C, \theta, L) = \theta \times \min(L(1 + r(L, C)), \tilde{C}). \quad (4)$$

When taking a short position in an asset, the investor needs to use cash as collateral to back up his promised cash flow. We use \tilde{K} to refer to the “credibly promised” cash flow to the investor who holds asset K . For example, if an investor promises a cash flow \tilde{A} and posts $V_u - \Delta$ cash as collateral, the credibly promised cash flow is

$$\begin{cases} V_u - \Delta & \text{if } \tilde{V} = V_u, \\ V_d + \tilde{U} & \text{if } \tilde{V} = V_d, \end{cases} \quad (5)$$

because when the realized value of \tilde{A} is greater than the collateral, this investor will default on his promise and his counterparty will get the collateral $V_u - \Delta$. That is, the real cash flow here is the one in equation (5), rather than the “artificial” promise \tilde{A} . Therefore, in our discussion below, we will use the credibly promised cash flow to denote the derivative. This implies that, without loss of generality, we can assume that the short seller of a derivative posts enough collateral and will not default.⁸ If an investor shorts one units of asset C , he needs to post $\max \tilde{C}$ cash collateral, where $\max \tilde{C}$ is the maximum value of the cash flow from asset C at time $t = 1$. Therefore, shorting θ units of asset C needs $\max \tilde{C} - P_C$ capital from the short seller and its payoff at time $t = 1$ is

$$W^-(C, \theta) = \theta(\max \tilde{C} - \tilde{C}), \quad \text{for } \theta \geq 0. \quad (6)$$

To finance a long position in asset C , an investor can choose any loan contracts (L, C) and

⁸For example, we can redefine the promise of \tilde{A} with $V_u - \Delta$ cash collateral as the promise of the cash flow in equation (5) with $V_u - \Delta$ cash collateral. There is obviously no default in this newly defined promise.

may choose multiple types of loan contracts to finance different parts of the position. We use $(C, \theta_{i,C}^+(L), L)$, for $C \in \{A, K\}$, $\theta_{i,C}^+(L) \geq 0$, and $L \in [0, P_C]$ to describe these long positions,⁹ where $\theta_{i,C}^+(L)$ is the number of units of asset C held by investor i and financed by (L, C) . We use $M_{i,C}^+(x)$ to denote investor i 's aggregate holding in asset C that has been financed by loan contracts (C, L) with $L \leq x$. Note that the relation between $\theta_{i,C}^+(\cdot)$ and $M_{i,C}^+(\cdot)$ is similar to that between a Probability Density Function and its corresponding Cumulative Distribution Function. It is easy to see that $M_{i,C}^+(x)$ is right continuous, weakly increasing and $M_{i,C}^+(0) = 0$. Similarly, we use $(C, \theta_{i,C}^*(L), L)$ to denote the levered long positions in asset C , which were financed by investor i ; and use $M_{i,C}^*(x)$ to denote the aggregate long position in asset C that is financed by investor i with loan contracts (C, L) with $L \leq x$. Clearly, $M_{i,C}^*(x)$ is also right continuous, weakly increasing and $M_{i,C}^*(0) = 0$. Finally, $\theta_{i,C}^-$ denotes the unit of asset C that are shorted by investor i and η_i denotes investor i 's investment in the riskless technology, with $\theta_{i,C}^- \geq 0$ and $\eta_i \geq 0$.

With these notations, we can denote investor i 's ($i = o, p$) wealth at time $t = 1$ as

$$W_i = \sum_{C=A,K} \left(\int_0^{P_C} W^+(C, \theta_{i,C}^+(L), L) dM_{i,C}^+(L) + W^-(C, \theta_{i,C}^-) + \int_0^{P_C} X(C, \theta_{i,C}^*(L), L) dM_{i,C}^*(L) \right) + \eta_i. \quad (7)$$

His objective is to choose his portfolio $(\theta_{i,C}^+(L), \theta_{i,C}^-, \theta_{i,C}^*(L), \eta_i)$ for $C = A, K$ and $L \in [0, P_C]$, to maximize his expected wealth at $t = 1$:

$$\max E_i(W_i) \quad (8)$$

$$s.t. \sum_{C=A,K} \left(\int_0^{P_C} \theta_{i,C}^+(L)(P_C - L) dM_{i,C}^+(L) + \int_0^{P_C} L \theta_{i,C}^*(L) dM_{i,C}^*(L) + \theta_{i,C}^- (\max \tilde{C} - P_C) \right) + \eta_i \leq e + \beta P_A. \quad (9)$$

where the left hand side of equation (9) is the total capital an investor allocates to long positions, lending, short positions, and the riskless technology; the right hand side is the investor's initial endowment.

Our focus is to analyze which derivative contract K and borrowing contract (L, C) will be adopted in equilibrium. Before introducing the notion of optimal innovation, however, we first analyze the equilibrium, taking the derivative contract as given.

⁹Since there is no penalty to default in this economy, no investor can borrow more than the collateral value. Hence, we don't need to consider the case of $L > P_C$.

Definition 1 *The equilibrium given derivative K is defined as the prices of assets A and K , (P_A, P_K) , investors' holdings, $(\theta_{i,C}^+(L), \theta_{i,C}^-, \theta_{i,C}^*(L), \eta_i)$ for $i \in \{o, p\}$, $C \in \{A, K\}$ and $L \in [0, P_C]$, and the interest rates $r(L, C)$ for all adopted loan contracts, such that for all investors, their holdings solve their optimization problem (8), and all markets clear:*

$$\sum_{i=o,p} \left(\int_0^{P_A} \theta_{i,A}^+(L) dM_{i,A}^+(L) + \theta_{i,A}^- \right) = \beta; \quad (10)$$

$$\sum_{i=o,p} \left(\int_0^{P_K} \theta_{i,K}^+(L) dM_{i,K}^+(L) + \theta_{i,K}^- \right) = 0; \quad (11)$$

and for $i \in \{o, p\}$, $j \neq i$, $C \in \{A, K\}$, and $L > \min \tilde{C}$:

$$\theta_{i,C}^*(L) = \theta_{j,C}^+(L). \quad (12)$$

Equations (10) and (11) state that the aggregate demand is β units for asset A and zero for asset K . Equation (12) implies that borrowing is equal to lending for all loan markets with $L > \min \tilde{C}$. Note that if $L \leq \min \tilde{C}$, this borrowing is riskless and can be done through the riskless technology, rather than borrowing from some other investors in the economy.¹⁰

Due to the disagreement on \tilde{V} , investors would like to speculate on its value at $t = 1$. It is natural to conjecture that, in equilibrium, investors would adopt a derivative contract, asset V , which is a claim to a cash flow \tilde{V} at $t=1$. Before we demonstrate that asset V will indeed be adopted, we first construct the equilibrium, taking the market for asset V as given. Our analysis next will focus on the case $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, where

$$\begin{aligned} \underline{\alpha} &\equiv \frac{\gamma - \gamma h_o}{\gamma + \beta h_o}, \\ \bar{\alpha} &\equiv 1 - \frac{h_p e + h_p \beta (V_u - V_d - \Delta)}{e + \beta [h_p (V_u - V_d) + V_d]}, \\ \gamma &\equiv \frac{h_o [e + \beta (V_d - \Delta)]}{h_o (V_u - V_d) + \Delta}. \end{aligned}$$

The equilibrium in other cases is uninteresting and is completely dominated by one group investors. For example, in the case $0 < \alpha_p < \underline{\alpha}$, there are so few pessimistic investors, so that the

¹⁰One interpretation is the following. The cash collateral in the economy is kept at a custodian bank, which can only invest the cash in riskless investment. So, if an investor has sufficient collateral to guarantee no default, he can borrow from this custodian bank at riskless interest rate.

equilibrium prices of assets A and V are completely determined by optimists' belief

$$P_A = \mathbf{E}_o[\tilde{A}],$$

$$P_V = \mathbf{E}_o[\tilde{V}].$$

Similarly, when $\alpha_p > \bar{\alpha}$, there are so many pessimistic investors, so that the equilibrium prices are completely determined by pessimists' belief. So our focus will be on the intermediate region $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, where the equilibrium is determined by the interaction between the two groups. To best illustrate the main insights in our model, we first analyze the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, where the value of α_1 is given by (26) in Appendix, and leave the analysis of the case $\alpha_1 \leq \alpha_p \leq \bar{\alpha}$ to Section 5.

Proposition 1 *In the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \frac{e\alpha_o + (\gamma + \beta)(V_d - \Delta)}{\gamma + \beta\alpha_p}, \quad (13)$$

$$P_V = \frac{\gamma\alpha_o}{\gamma + \beta\alpha_p}V_u + \frac{(\gamma + \beta)\alpha_p}{\gamma + \beta\alpha_p}V_d. \quad (14)$$

2. *A fraction $\frac{\beta}{\gamma + \beta}$ of type-o investors hold $(A, \frac{\gamma + \beta}{\alpha_o}, V_d - \Delta)$ and $r(V_d - \Delta, A) = 0$.*
3. *The rest of type-o investors hold $(V, \frac{e + \beta P_A}{P_V - V_d}, V_d)$ and $r(V_d, V) = 0$.*
4. *Type-p investors short $\frac{e + \beta P_A}{V_u - P_V}$ derivative V and posts V_u cash as collateral for each contract.*

In this equilibrium, optimistic investors are indifferent between holding a levered position in A and a levered position in the derivative V . They can borrow at the riskless interest rate if they can post enough collateral to guarantee no default. Alternatively, they can reach out to other investors to enter a debt contract, if both sides can agree on the collateral and the interest rate. For example, if an optimistic investor borrows $V_d - \Delta$ against each share of asset A as collateral, he can guarantee no default and so the interest rate is 0. If he wants to borrow more, however, he has to offer a higher interest rate to his lender to compensate the default risk. If

an investor from group p agrees to lend, this choice of lending has to be no worse than his outside option, which is taking a short position in V . In the case $\underline{\alpha} \leq \alpha_p < \alpha_1$, the optimistic investors cannot offer an interest rate that is high enough to attract pessimistic investors to lend to them. Similarly, optimistic investors cannot offer an interest rate that is attractive enough for pessimistic investors to finance their purchase of the derivative contract V . As shown in items 2 and 3 in the proposition, type- o investors are indifferent about those two strategies. A fraction $\beta/(\gamma + \beta)$ of them hold a levered position in asset A . Each of them holds $(\gamma + \beta)/\alpha_o$ units asset A . Using each unit as collateral, the investor borrows $V_d - \Delta$ at the riskless interest rate 0, since the collateral can guarantee no default. The rest of the optimistic investors take a levered long position in the derivative contract V .¹¹ Each of them holds $\frac{e + \beta P_A}{P_V - V_d}$ contract. Using each contract V as collateral, the investor borrows V_d and the interest rate is 0. Finally, pessimistic investors take a short position in the derivative contract V . Note that no pessimistic investors choose to short A , this is because, as we will see next, shorting the derivative V is more appealing.

These results highlight the main theme of this paper: one important motivation for financial innovation is to facilitate investors to set up their positions with a smaller amount of collateral. Each unit of asset A , or asset V , gives investors the same exposure to \tilde{V} , which they are interested in betting on. However, buying V requires less collateral since the investor can borrow more against V : An investor can borrow V_d against each unit of derivative V , but can only borrow $V_d - \Delta$ against each unit of asset A .

The reason that trading asset A needs more collateral is because of the “unwanted” risk from \tilde{U} . That is, the investor has to “waste” his collateral to cover the risk he is not interested in taking. This is unappealing even if the investor is risk neutral. The intuition suggests that since the derivative contract V completely carves out the unwanted cash flow, it is the “most appealing” financial innovation in this economy, as we formally analyze next.

¹¹Note that type- o investors are risk neutral and indifferent about the two strategies and so are also indifferent about any combination of the two strategies. Therefore, we can also interpret the result as “a fraction $\beta/(\gamma + \beta)$ of type- o investors’s wealth is invested in the levered position in A and the rest of their wealth in the levered position in V .”

3.6 Optimal Financial Innovation

Definition 2 *A financial innovation X (a claim to a cash flow \tilde{X} in \mathcal{H} at $t = 1$) is optimal if, in the presence of X , one introduced any other derivative contract K (a claim to a cash flow \tilde{K} in \mathcal{H} at $t = 1$), the market for K wouldn't generate any trading, unless \tilde{K} is perfectly correlated with \tilde{X} .*

Proposition 2 *The derivative contract V is an optimal financial innovation.*

Due to the disagreement, optimistic investors prefer to transfer their wealth at $t = 1$ to the up state and the pessimistic ones the down state. Derivative V is the most efficient instrument since it allows them to transfer all their wealth to the states they prefer. Alternative derivative contracts cannot achieve this goal. For example, let's consider a derivative contract that pays \tilde{A} at $t = 1$. That is, an investor with a long position in this derivative receives the same cash flow as that from the underlying asset A . As shown in Proposition 1, the investor can only borrow $V_d - \Delta$ against each unit of this asset as collateral. Therefore, this optimistic investor cannot completely transfer his wealth to the up state, since his wealth at the down state is always positive unless the realization of \tilde{A} happens to be $V_d - \Delta$. Similarly, the pessimist who shorts this derivative cannot transfer all his wealth to the down state. Therefore, trading V leads to Parato improvement since it enables both optimists and pessimists to transfer their wealth to the states they prefer. This explains why no investors short asset A in equilibrium. More generally, the above intuition implies that, in the presence of the market for V , any alternative derivative markets cannot generate any trading.

3.7 Implementation with Transaction Costs

Proposition 2 states that derivative contract V is an optimal security. It does not, however, pin down the unique contract in the economy. In fact, any linear transformation of asset V serves exactly the same function as asset V . For example, if asset X is a claim to a cash flow $\tilde{X} = a(\tilde{V} + b)$, where a and b are constants, it serves the same economic function as asset V . To see this, we note that a can be normalized to 1 by redefining the size of each unit. So, we

only need to consider the case $\tilde{X} = \tilde{V} + b$. The difference between assets X and V is that asset X pays an extra constant cash flow b . Not surprisingly, if we introduce asset X into the economy, its price would be $P_X = P_V + b$. To take a long position in asset X , an investor can get a loan contract $(V_d + b, X)$ and the interest rate is 0. Therefore, the payoff to the position $(X, 1, V_d + b)$ is identical to the payoff to $(V, 1, V_d)$. That is, asset X serves exactly the same economic function as asset V for any b .

In the next, however, we will illustrate that any infinitesimal transaction cost can pin down the unique optimal contract. Imagine, for this section only, that there is a small transaction cost for transferring funds from one account to another. Specifically, the cost for transferring M dollars from one investor to another is kM dollars and the sender and receiver each pay $kM/2$, where k is positive constant. For simplicity, we assume that k is infinitesimal. Therefore, at $t = 0$, investors are facing essentially the same problem as analyzed before and the equilibrium prices are the same as those in Proposition 1. Moreover, investors can minimize their transaction costs by “fine-tuning” the derivative contract, as shown in the following proposition.

Proposition 3 *In the presence of the transaction cost described above, the optimal financial innovation is determined by a unique value of b , such that $P_X = 0$.*

This proposition illustrates the appealing characteristic of derivatives with a zero initial price, a common feature in many derivatives in practice e.g., futures and swaps. They are so-called “unfunded” securities, with the name highlighting the fact that investors can establish their positions without paying at inception and only need to post collateral to cover the daily mark-to-the-market movements.

It is very intuitive to see why unfunded securities are appealing. Suppose we had chosen b such that the contract’s initial value is not zero. Then the cash flows from trading this security can be decomposed into two components. The first component is the cash flows from trading a corresponding unfunded derivative. The second component is the following: Since the initial price of the derivative is not zero, one party gets paid at $t = 0$, but then he needs to pay this amount back at $t = 1$. Note that while the first component serves the economic function by facilitating the speculation among investors, the second one is completely redundant. Making the

derivative unfunded avoids this potential “round trip” in fund transfer. To the extent that there is an infinitesimal cost of transferring funds, the unfunded security would be strictly preferred.

4 Violation of the Law of One Price

As noted earlier, holding asset A is inefficient due to its higher collateral requirement. In equilibrium, therefore, to induce an investor to hold it, there has to be a price discount relative to the derivative V . This can potentially lead to the violation of the law of one price, that is, the equilibrium price of an asset can be different from the price of its replicating portfolio.

Before we proceed with our analysis, it is helpful to first describe the empirical motivation. One example is the so-called corporate-bond-CDS basis, the difference between the CDS spread and the corresponding corporate bond yield spread. As noted in Mitchell and Pulvino (2010) and Garleanu and Pedersen (2011), CDS spreads tend to be lower than the corresponding corporate bond yield spreads, although both are measures of the underlying firm’s credit risk and the no arbitrage relation implies that the difference between the two should be near zero. In other words, a corporate bond can be decomposed into a short position in a CDS contract on the bond issuer plus a Treasury bond. The empirical evidence suggests that the price of the corporate bond is often lower than the price of the portfolio of the CDS and Treasury bond.

Keep this example in mind, let’s now examine the prices in our model. Note that asset A can be decomposed into assets V and U , where asset U is a claim to a cash flow \tilde{U} at $t = 1$. That is, to analyze the violation of the law of one price, we need to compare P_A with $P_V + P_U$, where P_U is the price of asset U . Note that P_A and P_V have been determined in Proposition 1. How is P_U determined? To see this, it is helpful to map our model to the earlier example. Assets A , V and U in our model correspond to the corporate bond, CDS, and Treasury bond, respectively. So, how is Treasury bond (asset U) price determined? It is obviously jointly determined by a large number of investors, many of whom are not involved in the corporate bond market at all. Therefore, a natural way to think of asset U is to assume that there is another market (the Treasury market in our example), in which a large number of investors trade asset U , and these investors (e.g., sovereign funds, repo trading desks at investment banks) do not trade assets A

and V . If the investors in asset U market are all risk neutral and don't have funding constraints (e.g., sovereign funds) we have $P_U = 0$. On the other hand, if those investors are risk averse or face funding constraints, P_U is negative. Finally, if investors are attracted by some special features of asset U , its price can be positive. This can happen, for example, during fly-to-quality in crises, or more generally when investors treat Treasury securities as money and value their convenience yield (Krishnamurthy and Vissing-Jorgensen (2010)). Note that the both type- o and type- p investors in our model have access to this asset U market, but would choose not to trade U if P_U is close to 0. In this case, the equilibrium prices of A and V are still the same as those in Proposition 1.

4.1 Model Implications

We use B to denote the basis, the price difference between asset A and its replicating portfolio: $B \equiv P_V + P_U - P_A$. Naturally, the basis can be decomposed into two components:

$$B = S + P_U, \tag{15}$$

where $S \equiv P_V - P_A$. The first component S reflects the value of saving collateral. Even though all investors are risk neutral, they still value A less than V , because V allows investors to bet with less collateral. The second component is derived from the fact that asset U (e.g., Treasury) is traded in a much larger market, and is probably more liquid and has certain specialness, as discussed in Krishnamurthy and Vissing-Jorgensen (2010). It is important to note that the second component reflects how much value investors assign to the liquidity of the Treasury market, and so affects the basis for all assets. For example, fly-to-quality during crises increases P_U and so the basis for all assets equally. In contrast, the first component, S , depends on the characteristics of asset A and hence also has both cross-sectional and time series implications on basis, as characterized in the following proposition.

Proposition 4 *The price spread S is positive and has the following properties.*

- (1) S increases when there is less cash in the economy: $\frac{\partial S}{\partial e} < 0$;
- (2) S increases when asset A has more unwanted risk: $\frac{\partial S}{\partial \Delta} > 0$;

(3) The impact in (1) is stronger when there is more unwanted risk: $\frac{\partial^2 S}{\partial e \partial \Delta} < 0$;

(4) Suppose an outside investor has to sell β^* units of asset A to the investors in this economy.

The spread increases: $\frac{\partial S}{\partial \beta^*} > 0$, and this impact is stronger when there is more unwanted risk: $\frac{\partial^2 S}{\partial \beta^* \partial \Delta} > 0$;

(5) S increases when there is more trading need in the economy: $\frac{\partial S}{\partial h_0} > 0$, $\frac{\partial S}{\partial \alpha_o} < 0$.

Result (1) shows that this spread increases when investors have less cash, i.e., when there is less funding liquidity in the market. This is because saving collateral becomes more valuable when investors have less cash but need leverage. Similarly, Result (2) says that the spread is larger if the unwanted portion of the cash flow, \tilde{U} , is more volatile (i.e., Δ is larger). This is because the risk in \tilde{U} determines how much collateral can be saved by trading V . The larger the risk in \tilde{U} , the more collateral can be saved by trading V , leading to a larger price spread. Related with these two results, Result (3) shows that when the funding liquidity in the economy tightens (i.e., e decreases) the spread increases more for assets with more volatile unwanted cash flow (i.e., larger Δ).

Result (4) is about the impact from supply shocks. If a large investor has to liquidate his positions in asset A at $t = 0$, what is the impact of this supply shock to equilibrium prices? Clearly, the prices of both A and V will drop. Result (4) shows that the price of A drops more, i.e., the supply shock increases the spread. This is because it takes more capital to absorb A than to absorb V , implying that the price of A is more sensitive to supply shocks. Similarly, the spread is more sensitive to supply shocks if the asset has a larger unwanted risk.

Finally, the spread increases with investors' trading motive. This is because the spread reflects the value investors assign to saving collateral in their trading. In the model, the trading motive increases with h_0 . This is because, holding everything else constant, an increase in h_0 (i.e., the optimists become even more optimistic) increases the disagreement and so the trading motive. Similarly, the trading motive increases in α_p . This is because, in the current case, the investors are predominately optimists. An increase in pessimists' population size α_p makes it more balanced between the optimists and pessimists, leading to a stronger trading motive.

4.2 Existing Evidence and Further Testable Predictions

The above implications shed light on the empirical evidence from a number of studies. The previously-mentioned corporate-bond-CDS basis arises naturally in our model. Suppose an investor, say a hedge fund, wants to take an exposure on a corporate bond. He can either buy this bond on margin (i.e., using the bond as collateral to finance the purchase) or he can simply short a CDS contract on this firm. Intuitively, to establish the same exposure to the default risk of the firm, the corporate bond position takes more collateral because it also has embedded interest rate risk. In other words, if the corporate-bond-CDS basis were zero, shorting CDS would be more desirable to the investor. In equilibrium, therefore, the CDS rate is lower, leading to a positive corporate-bond-CDS basis.

Moreover, consistent with Result (1) of Proposition 4, the CDS-corporate-bond basis increased substantially during the recent financial crisis, when the funding liquidity was probably tight for most investors. It is possible that fly-to-quality during the crisis disproportionately increased the price of Treasury bonds (P_U in our model) and so contributed to part of the observed increase in basis. However, this interpretation cannot account for the cross-sectional variation in basis. As documented in Mitchell and Pulvino (2010), during the crisis, the corporate-bond-CDS basis tends to be larger for junk bonds than for investment grade bonds. This evidence is consistent with Result (2), if one takes the interpretation junk bonds have more non-default-related risks (e.g., liquidity risk). Moreover, with the financial crisis unfolding, the basis for junk bonds increases more than that for investment grade ones, consistent with Result (3).

Another example is the discrepancy between the expected inflation implied in the inflation swaps market and that implied in the TIPS market. Fleckenstein, Longstaff and Lustig (2010) find that the price of TIPS is consistently lower than the price of its replicating portfolio that consists of inflation swaps and nominal Treasury bonds. They suggest that part of this phenomenon can be attributed to “margins, haircuts, and other collateral-related frictions.” Our model formalizes this intuition. If an investor decides to hedge inflation, or speculate that inflation will go up, he can buy TIPS, or take a long position in inflation swaps. Note that, relative to the former strategy, the latter needs less collateral to establish the same exposure to inflation.

This is because the interest rate risk embedded in TIPS increases the collateral requirement for trading TIPS. As noted in Fleckenstein, Longstaff, and Lustig (2010), the haircuts for TIPS is around 3% in their sample. That is, the collateral requirement for purchasing TIPS is around 3%. In contrast, an inflation swap is an unfunded security. To take a long position, one just needs to post collateral to cover the daily movements in inflation swap rates. So the collateral requirement is much smaller. Hence, our model implies that other things equal, investors would prefer the long position in inflation swap, leading to a swap rate that is higher than what is implied by TIPS and nominal Treasury bonds. Moreover, Fleckenstein, Longstaff and Lustig (2010) also show that, consistent with Result (1), this price discrepancy between inflation swaps and TIPS increased dramatically during the financial crisis, when funding liquidity was perhaps tight for most investors.

Proposition 4 also offers a number of testable predictions. Result (2) suggests that other things equal, the corporate-bond-CDS basis and the TIPS-inflation-swap basis should be larger for bonds with higher unwanted risks. This implies, for example, that the basis should be larger for corporate bonds or TIPS with longer maturities, or when the riskless interest rate volatility is higher. Moreover, Result (4) implies that the basis should increase when there is a positive supply shock to the underlying asset (e.g., when a large institution is forced to liquidate its corporate bonds or TIPS). This supply shock impact should be stronger for longer maturities, or when the riskless interest rate volatility is larger. Finally, Result (5) implies that, all else being equal, the basis should be larger when trading motive is stronger.

Note that as shown in (15), the observed basis has two components. The second component (i.e., the specialness of nominal Treasury bonds) might have also contributed to the observed basis. For example, fly-to-quality during the recent financial crisis might have contributed to the large increases in the corporate-bond-CDS basis and TIPS-inflation-swap basis. However, this force cannot explain the observed cross-sectional variations. There are also other factors that may have contributed to the observed corporate-bond-CDS basis. One example is counterparty risk. As dealers' default probability increases during the crisis, the CDS contracts they underwrite become less valuable, leading to a lower CDS spread and so a higher corporate-bond-CDS

basis. This is certainly possible. Arora, Gandhi, and Longstaff (2010) find that counterparty risk is indeed priced in the CDS market. However, they note that perhaps due to the common practice of full collateralization of swap liabilities, the impact on CDS spread is very small. Moreover, it is not clear whether counterparty risks increase or decrease the TIPS-inflation-swaps basis. To the extent that the concern of counterparty risks reduces the value of inflation hedge offered by weakened institutions, this would decrease the inflation swap rates and so decreases the basis, opposite to the evidence.

5 Other cases

The analysis so far is focused on the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$. This section presents the results from other cases. Note that there is no default risk in equilibrium in the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$. For example, using one share of asset A as collateral, an investor chooses to borrow $V_d - \Delta$. Hence, even in the worst case, the value of the collateral is enough to pay back the debt. The investor has the choice to borrow more against the collateral. But, in the case $\underline{\alpha} \leq \alpha_p \leq \alpha_1$, the lender would charge an interest rate that is too high, so that the borrower prefers to borrow only $V_d - \Delta$ to get the riskless interest rate. In the case of $\alpha_1 \leq \alpha_p \leq \bar{\alpha}$, however, lenders can offer a rate that is also acceptable to the borrowers if they choose to borrow more. Hence, in equilibrium, some of the borrowing has default risk. Depending on the relative sizes of the two group of investors, the equilibrium can be characterized by two subcases. In the first case, $\alpha_1 \leq \alpha_p < \alpha_2$, only a fraction of the asset- A -backed debts has default risk; while in the other case, $\alpha_2 \leq \alpha_p \leq \bar{\alpha}$, all asset- A -backed debts have default risk, where the expression for α_2 is given by equations (41) in the Appendix.

Proposition 5 *In the case $\alpha_1 \leq \alpha_p < \alpha_2$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \frac{1}{z^* + 1} V_u + \frac{z^*}{z^* + 1} V_d - \left(1 - \frac{1}{h_o} \frac{1}{z^* + 1}\right) \Delta, \quad (16)$$

$$P_V = \frac{1}{z^* + 1} V_u + \frac{z^*}{z^* + 1} V_d, \quad (17)$$

where $z^* \equiv \frac{(1-\alpha_2)(\gamma+\beta)}{\alpha_2\gamma}$.

2. *Optimistic investors are indifferent about the following three strategies.*

- A measure x_o^* of them hold a position $(V, \frac{e+\beta P_A}{P_V-V_d}, V_d)$ and $r(V_d, V) = 0$, where x_o^* is given by equation (36).
- A measure y_o^* of them hold a position $(A, \frac{e+\beta P_A}{P_V-V_d+\Delta}, V_d - \Delta)$, and $r(V_d - \Delta, A) = 0$, where y_o^* is given by equation (40).
- The rest of them hold a position $(A, \frac{e+\beta P_A}{P_V-L^*}, L^*)$, $r(L^*, A)$ is positive and given by (42), where L^* is given by (38).

3. *Pessimistic investors are indifferent about the following two strategies.*

- A measure x_p^* of them short $\frac{x_o^*(e+\beta P_A)}{x_p^*(P_V-V_d)}$ contract V , where $x_p^* = z^* x_o^*$.
- The rest of them lend their wealth, $e + \beta P_A$, to optimistic investors. The lending contract is (L^*, A) and the interest rate is given by (42).

The equilibrium in this case is similar to that analyzed in Proposition 1. The only difference is that some optimists can now take more leverage, but need to pay a higher interest rate to compensate the lenders for the credit risk in the loan. More precisely, optimists are indifferent about the three strategies: A measure x_o^* of optimists choose to take a levered position in derivative V . The rest of the optimists take long positions in asset A , but they have two different ways to finance their positions. A measure y_o^* of them choose to borrow less, so that they face the riskless interest rate. The rest of them, however, choose to borrow more and face a higher interest rate. The bigger loan enables them to have a larger position and this extra expected profit compensates the higher interest they face.

Another new feature in this case is that as shown in equations (16) and (17), the prices of assets A and V are independent of α_p . In contrast, in the equilibrium in Proposition 1, both P_A and P_V decrease in α_p . This is intuitive. More pessimists take short positions in V , pushing down P_V . This attracts more optimists from A to V , and hence pushes down P_A as well. In the case of Proposition 5, however, there is another force. With the increase of α_p , more pessimists

choose to lend to optimists. This enables optimists to borrow to invest more, pushing up P_A and P_V . These two forces exactly offset each other in the case in Proposition 5, so that P_A and P_V do not depend on α_p .

Proposition 6 *In the case $\alpha_2 \leq \alpha_p \leq \bar{\alpha}$, the equilibrium is characterized as follows:*

1. *The prices of assets A and V are given by*

$$P_A = \left(\frac{1}{x_o^{**} + x_p^{**}} - 1 \right) \frac{e}{\beta}, \quad (18)$$

$$P_V = \frac{x_o^{**}}{x_o^{**} + x_p^{**}} V_u + \frac{x_p^{**}}{x_o^{**} + x_p^{**}} V_d, \quad (19)$$

where x_o^{**} and x_p^{**} are given by (44) and (45).

2. *Optimistic investors are indifferent about the following two strategies.*

- *A measure x_o^{**} of them hold a position $(V, \frac{e+\beta P_A}{P_V-V_d}, V_d)$ and $r(V_d, V)=0$.*
- *The rest of them hold a position $(A, \frac{e+\beta P_A}{P_V-L^{**}}, L^{**})$, and $r(L^{**}, A)$ is positive and given by (47), where L^{**} is given by (43).*

3. *Pessimistic investors are indifferent about the following two strategies.*

- *A measure x_p^{**} of them shorts $\frac{x_o^{**}(e+\beta P_A)}{x_p^{**}(P_V-V_d)}$ contract V.*
- *The rest of them lend all their wealth, $e + \beta P_A$, to optimistic investors. The lending contract is (L^{**}, A) and the interest rate is given by (47).*

Similar to the previous case, some of the optimists' borrowing has default risk. One difference is that in this case, when optimists use A as collateral to borrow, they all prefer to borrow more than $V_d - \Delta$ and pay a positive interest rate.

Putting together all three cases in Propositions 1, 5 and 6, we obtain the plots in Figure 1. The upper panel plots P_A and P_V against α_p . At $\alpha_p = \underline{\alpha}$, both P_A and P_V are completely pinned down by optimists' expectation $P_A = \mathbf{E}_o[\tilde{A}]$ and $P_V = \mathbf{E}_o[\tilde{V}]$. As the population size of pessimists increases, both prices decrease in the case of $\underline{\alpha} \leq \alpha_p < \alpha_1$. In the case of

$\alpha_1 \leq \alpha_p < \alpha_2$, however, both prices stay constant while α_p changes, as shown in Proposition 5. Finally, in the region $\alpha_2 \leq \alpha_p < \bar{\alpha}$, when α_p increases, P_V decreases but P_A increases. The reason is that two forces arise when pessimists' population size increases. First, their larger short positions in V pushes down its price and this attracts more optimists to take long positions in V , reducing the number of optimists holding A . On the other hand, more pessimists compete to lend to optimists, and push down the interest rate on asset- A -backed debts, giving optimists more purchasing power. This second impact can dominate and so P_A increases with α_p .

Figure 1: Asset Prices and Interest Rates.

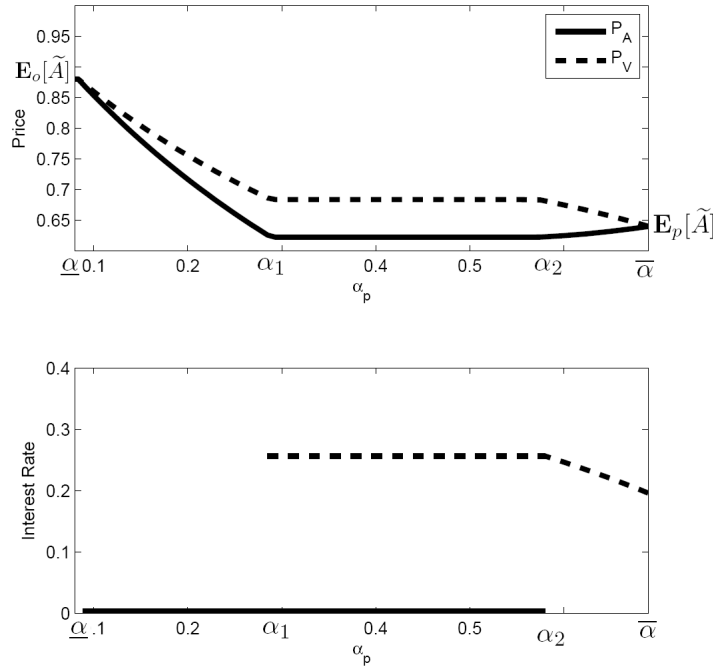


Figure 1: The upper panel plots P_A and P_V against α_p . The lower panel plots the notional interest rates on loans backed by asset A . Parameter values: $h_p = 0.4, h_o = 0.8, V_u = 1, V_d = 0.4, \beta = 1, e = 0.2, \Delta = 0.15$, and \tilde{U} is uniformly distributed.

One can see the above intuition more clearly by examining the interest rates. Across all three cases in the region $\underline{\alpha} \leq \alpha_p \leq \bar{\alpha}$, all asset- V -backed debts are riskless and have a zero interest rate. In contrast, the credit risk of asset- A -backed debts varies across cases. As shown in the lower panel of Figure 1, asset- A -backed debts are riskless and have a zero interest rate in the case of $\underline{\alpha} \leq \alpha_p < \alpha_1$. In the case of $\alpha_1 \leq \alpha_p < \alpha_2$, however, some of the asset- A -backed debts

are riskless and have a zero interest rate while the rest have default risk and have a positive interest rate, which stays a constant throughout the region. Finally, in the case of $\alpha_2 \leq \alpha_p < \bar{\alpha}$, all asset- A -backed debts have default risk. Note that the interest rate drops when α_p increases, indicating that when more pessimists compete to lend, they push down the interest rate. This benefits the borrower (i.e. the optimists) and can even lead to the result in the upper panel that P_A increases with α_p .

6 The Impact of Financial Innovation

How does financial innovation affect the economy? To analyze this, we compare the equilibria across two economies. The first is the above economy with assets A and V . As a comparison, the second economy does not have the market for V and is otherwise identical to the first economy. The analysis of the second economy is similar to that in previous sections and we leave the details to the Appendix. In the following, we summarize the impact of financial innovation by comparing the equilibria across these two economies.

Proposition 7 *Introducing the market for V may increase, decrease, or have no impact on the price of asset A .*

The intuition is as follows. The derivative contract V is efficient in facilitating investors' bets. On the one hand, optimists prefer to buy asset V , rather than the underlying asset A . This puts downward pressure on the price of asset A . On the other hand, pessimists are also attracted to shorting V , away from shorting A . This increases the price of A . The overall impact on asset A is mixed, and determined by the tradeoff between these two forces.

Interestingly, with the presence of asset V , the price of asset A can be even lower than the pessimists' expected value $\mathbf{E}_p[\tilde{A}]$. Note that in the economy without asset V , the price of asset A is always between $\mathbf{E}_p[\tilde{A}]$ and $\mathbf{E}_o[\tilde{A}]$, the expected values of the two groups of investors. This is natural. If the price of A were less than $\mathbf{E}_p[\tilde{A}]$, for instance, both investors would want to buy it, which would have pushed up the price. In the presence of the derivative V , however, Figure 1 shows that when α_p is large, the price of asset A is even lower than the pessimist's expected

value, $\mathbf{E}_p[\tilde{A}]$. Although pessimists find it profitable to buy asset A , they choose not to do so because they find trading asset V even more profitable.

In our previous discussion, we mostly take the heterogeneous belief interpretation. This makes it harder to examine the welfare implications because it is unclear which belief should be used when calculating investors' welfare.¹² As noted in Section 3.1, one can simply adopt the hedging interpretation. Investors' welfare under the hedging interpretation is mathematically identical to their subjective expected utility under the old heterogeneous-belief interpretation, and is reported in the following.

Proposition 8 *The introduction of V has a mixed impact on investors' welfare.*

Intuitively, the derivative contract V helps investors to transfer their wealth to the states they prefer, and so improves their welfare. However, the introduction of V also affects asset prices in the economy and so affects investors' welfare. For example, if the innovation increases the price of asset A , it decreases the expected utility of optimists since they now have to buy the asset at a higher price. On the other hand, the innovation increases the optimists' welfare if it decreases the price of asset A . As a result, financial innovation's mixed impact of the price of asset A , as shown in Proposition 7, translates into the mixed impact on investors' welfare.

More broadly, our analysis suggests that financial innovation allows investors to effectively take on more leverage. A more comprehensive welfare analysis should take into account the following two factors. On the one hand, innovation leads to more speculation, which has both positive and negative consequences. On the other hand, it also makes hedging cheaper and more effective. It is of course an empirical question to determine which effect is more important. We argue that speculation perhaps also played a significant role in driving financial innovation and trading in derivative markets. For example, the discussions on the Greek sovereign debt crisis suggest a strong concern about the speculation enabled by financial innovation. For instance, the efforts to push for a "voluntary" writedown are often attributed to the concern that a default

¹²Brunnermeier and Xiong (2011) proposes a solution to welfare analysis with heterogeneous beliefs for some cases.

would trigger large payments from CDS contracts and lead to chaos.¹³ This concern seems less consistent with the premise that most of the CDS positions were established for hedging purpose.

7 General Discussions

7.1 Common Theme Behind Various Innovations

This collateral view highlights the common theme behind a variety of financial innovations, despite their strikingly different appearances. For example, many successful derivative contracts are unfunded and allow investors to take on large positions with very little collateral. In addition to the previously mentioned swaps, futures is another example. A futures contract allows investors to take an exposure to the fluctuations of the price of a certain asset without the physical process of buying or selling the asset. This is especially important for commodity futures where transactions are costly and time-consuming. In other words, similar to the case for swaps, futures contracts save collateral by isolating the variables that investors want to bet on, i.e., the underlying asset prices. Of course, another feature, not captured in our model, is that futures contracts help to reduce collateral requirements by increasing the speed of transactions. Knowing that they can liquidate their clients' positions quickly to avoid losses, dealers do not need to demand a high collateral level.

This collateral view of innovation is not restricted to the invention of new securities. It also sheds light on the evolution of a legal practice, the *de facto* superseniority of derivatives and repos. Although derivatives and repos are not supersenior in a strict statutory sense, it has been a common practice in the U.S.: When a company goes bankrupt, its repo and derivative counterparties can simply seize the collateral posted in the transactions up to the amount the company owes them, instead of going through the lengthy and costly bankruptcy procedure. This exceptional treatment accorded derivatives and repos in bankruptcy is quite recent and has been evolving over time. In the context of our analysis, this practice can be viewed as carving unwanted cashflows out of derivative and repo transactions: Suppose an investor enters an interest rate swap and his goal is to hedge or speculate on interest rate risk, rather than

¹³See, e.g., Volunteers Wanted, *Economist*, January 21, 2012.

taking an exposure to his counterparty's credit risk. Without this superseniority, even if his counterparty posts a large amount of collateral, the investor would still not feel safe since he would have to go through automatic stay when his counterparty defaults. With superseniority, however, the investor can immediately seize the collateral upon default, and so can be better protected even by a smaller amount of collateral. In other words, superseniority separates the counterparty's assets into two parts, the collateral posted to the swap transaction and other assets. Given the investor's purpose, his counterparty's assets, apart from the collateral posted for the interest rate swap, are unwanted cashflows. The collateral efficiency is achieved when these unwanted cashflows are carved out of the swap transaction by superseniority.

Financial innovation may also take the form of new legal entities. For example, special purpose vehicles (SPVs) have become prevalent in recent decades with the rise of securitization. In the context of our analysis, we can view creating an SPV as, again, carving out unwanted cash flows (i.e., the firm's assets other than those allocated to the SPV). This interpretation is similar to the theory proposed in Gorton and Souleles (2006), which emphasizes the benefit of making SPVs bankruptcy remote to avoid bankruptcy cost.

The collateral friction in our model arises from the imperfection in cross-netting. In practice, it is becoming increasingly possible to have limited cross-netting. For example, on December 12 2006, the Securities and Exchange Commission (SEC) approved a rule change which made limited cross netting available to some investors in the exchange-traded options market. Another example is the efforts from brokers and hedge funds that attempt to get around the regulation-induced margin requirements (see, e.g., Brunnermeier and Pedersen (2009)). One can view the continuing efforts by regulators and market participants in modifying the margin procedure as one form of collateral-motivated financial innovation. Their goal is simply to satisfy the demand from market participants to alleviate their collateral constraints.

7.2 Open Issues

In practice, the collateral motivations are likely to play many more roles than what is captured in the above model. This section briefly discusses alternative roles played by collateral in fi-

nancial innovations. For example, our model assumes that all investors can easily handle the procedure of posting collateral. In practice, it seems natural to expect some derivative users may find it more costly to deal with posting collateral and daily marking-to-the-market. For example, corporations may find it onerous to deal with these issues and the induced cash flow uncertainty. As noted by Ross (2011), many OTC derivatives are very similar to those traded on exchanges, and the volume in the OTC markets is much larger. One conjecture is that the collateral flexibility for OTC contracts might be the key. According the data from The Federal Reserve Board, an average bank or broker only demands 0.1% collateral for its derivative exposures to corporation counterparties at the end of 2010. In contrast, it demands 72% collateral for its exposures to hedge funds and 45% for exposures to other banks or brokers. This tremendous variations simply highlight many unanswered questions. For example, what is the role of collateral in this financial system with highly heterogeneous institutions? What is the optimal innovation in this more complex world? Analysis of these issues is likely to shed light on the role of financial market in the overall economy and related policy issues.

Another open issue is the impact of financial innovation on market liquidity. For example, Dang, Gorton, and Homstrom (2011) show that information-insensitive securities discourage information production, which avoids adverse selection and hence is beneficial for liquidity provision. That is, “ignorance is a bliss for liquidity.” Our analysis, however, shows that financial innovation helps investors to take larger positions. This naturally encourages investors to produce more information, and so may jeopardize some of the liquidity benefits from information-insensitive securities. We leave these issues to a separate studies.

8 Conclusion

This paper proposes a collateral view of financial innovation. Many successful financial innovations, despite their strikingly different appearances, share the common motive of reducing collateral requirements to facilitate trading. We illustrate this insight in an equilibrium model in which both the financial market structure and collateral requirements are endogenously determined. We show that investors can save collateral in their trades by taking positions in

securities that carve out all “unwanted” cash flows. This financial innovation is “optimal” in the sense that its existence would drive out other derivative markets: if one introduced any other derivatives, those markets would not generate any trading. The model not only has a number of asset-pricing implications that are broadly consistent with existing empirical evidence, but also leads to some new testable predictions.

Appendix

Proof of Proposition 1

We first conjecture that the equilibrium is as follows: $x_o \in [0, \alpha_o)$ optimists invest all their wealth in a levered long position in asset V and the remaining, $\alpha_o - x_o$, invest all their wealth in a levered long position in asset A and all pessimists short asset V and use all their wealth in cash as collateral. Moreover, using each share of asset A as collateral, an investor can borrow $V_d - \Delta$, and the interest rate is 0. Using each contract V as collateral, the investor can borrow V_d and the interest rate is 0. To short each contract, the investor needs to put the V_u cash as collateral. We then derive the market clearing prices under this conjecture. Finally, we verify that the above conjecture is indeed sustained in equilibrium.

Note that in order to have a long position in asset V , the investor has to have $P_V - V_d$ capital since he can use the asset as collateral to borrow V_d . So the aggregate demand from x_o optimists is $x_o \frac{e + \beta P_A}{P_V - V_d}$. Similarly, pessimists' aggregate short position in asset V is $\alpha_p \frac{e + \beta P_A}{V_u - P_V}$. So the market clearing condition in the market for asset V is:

$$x_o \frac{e + \beta P_A}{P_V - V_d} = \alpha_p \frac{e + \beta P_A}{V_u - P_V}. \quad (20)$$

Similarly, the market clearing condition in the market for asset A is:

$$(\alpha_o - x_o) \frac{e + \beta P_A}{P_A - (V_d - \Delta)} = \beta. \quad (21)$$

Moreover, the expected utility for an optimist to borrow V_d to hold one share of asset V is $\mathbf{E}_o[\tilde{V}] - V_d$. So the expected utility from investing one dollar in this levered position in asset V is $\frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d}$. Similarly, the expected utility from investing one dollar in the levered position in asset A is $\frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A - (V_d - \Delta)}$. An optimist should be indifferent between these two strategies:

$$\frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d} = \frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A - (V_d - \Delta)}. \quad (22)$$

Similarly, for a pessimist, the expected utility from one dollar investment in shoring asset V is

$$\frac{V_u - \mathbf{E}_p[\tilde{V}]}{V_u - P_V}. \quad (23)$$

From (20)–(22), we obtain (13), (14) and $x_o = \beta/(\beta + \gamma)$.

We now turn to verify that this is an equilibrium by showing that no investor has incentive to deviate. Specifically, we need to verify the following:

- (a) No investor prefers to invest in the riskless technology.
- (b) Investor p prefers to short V rather than shorting A .
- (c) Investor o prefers to finance his long position in A by the borrowing contract $(V_d - \Delta, A)$.
- (d) Investor o prefers to finance his long position in V by the borrowing contract (Y_d, V) .

It is easy to verify that $\mathbf{E}_o[\tilde{V}] < P_V < \mathbf{E}_o[\tilde{V}]$. Therefore trading V strictly dominates investing in the riskless technology, implying (a). It is also straightforward to verify (b) by directly calculating the expected utility from shorting V and shorting A .

Clearly, investor o prefers the loan contract $(Y_d - \Delta, A)$ over (L, A) with $L < Y_d - \Delta$. This is because both loan contracts have zero interest rate but investor o 's expected return for the investment in asset A is positive. Hence, investor prefers the contract that allows him to borrow more. Hence, to prove (c), we need to verify that investor o prefers the loan contract $(Y_d - \Delta, A)$ over (L, A) with $L > Y_d - \Delta$. Note that with $L > Y_d - \Delta$ the loan contract has default risk. Hence the borrower has to compensate the lender by offering a higher interest rate. We need to check whether a type- o investor can borrow from a type- p investor with the contract (L, A) . Equations (22) and (23) imply that for the contract (L, A) to be preferred by both types of investors, the following two inequalities have to hold

$$\frac{\mathbf{E}_p \left[\min \left\{ \tilde{A}, L(1+r) \right\} \right]}{L} \geq \frac{V_u - \mathbf{E}_p[\tilde{V}]}{V_u - P_V}, \quad (24)$$

$$\frac{\mathbf{E}_o \left[\max \left\{ \tilde{A} - L(1+r), 0 \right\} \right]}{P_A - L} \geq \frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d}, \quad (25)$$

where r is the notional interest rate in the loan contract, the left hand side of (24) is the investor p 's expected return from the lending, and the left hand side of (25) is investor o 's expected return from the position $(A, 1, L)$.

By changing the inequalities in (24) and (25) into equalities, we obtain an equation system of L and r . We show in the online appendix that there exists a unique value α^* , $0 < \alpha^* < 1$, such that if $\alpha_p = \alpha^*$ there is a unique solution for this equation system. We define

$$\alpha_1 \equiv \alpha^*. \tag{26}$$

The appendix also shows that if $\underline{\alpha} \leq \alpha_p < \alpha_1$, inequalities (24) and (25) cannot hold simultaneously for any values of L and r . Therefore, this verifies (c). The proof for (d) is similar.

Proof of Proposition 2

In the following, we sketch the intuition of the proof. The details of the proof are left to the online appendix.

Step 1. In the case of $\beta = 0$, the resulting equilibrium is Pareto efficient. Investors have transferred all their $t = 1$ wealth to the states they prefer. It is easy to see that if the derivative K is not perfectly correlated with V , investors would strictly prefer not to trade it.

Step 2. Let's now consider the case of $\beta > 0$. In the presence of assets V and A , there will be 3 groups of investors in equilibrium. Group 1 long V and have an expected utility of J_1 , group 2 short V and have an expected utility of J_2 . Investors in other groups (e.g., the group that longs A) will be indifferent between their strategy and one of the two strategies adopted by groups 1 and 2.

Step 3. Let's now create a hypothetical economy, which is populated by groups 1 and 2 only. Their endowments are the same amount as those in the original equilibrium, but all in cash. Suppose these investors can trade asset V . It is easy to verify that in this hypothetical economy, the equilibrium is Pareto efficient and group i ($i = 1, 2$) investors' expected utility is still J_i . Moreover, the result in Step 1 implies that if the derivative K is not perfectly correlated with V , investors would strictly prefer not to trade it.

Step 4. Suppose there is a derivative \tilde{K} that is not perfectly correlated with \tilde{V} and generate some trades in the original economy, that leads to Pareto improvement. Then it must be the case that we can find investors from groups 1 and 2 such that trading K leads to Pareto improvement

among them. This implies that if we introduce K into the hypothetical economy in step 3, it would also generate trading there. This contradicts the conclusion in Step 3.

Proof of Proposition 3

Suppose X is an unfunded security. The flows of funds for those investors who trade X are the following: There is no need to transfer funds across investors at $t = 0$ since $P_X = 0$. At $t = 1$, the dealer just need to transfer all short sellers' wealth to long side if V_u is realized, or all the long side's wealth to the short side if V_d is realized. Now, suppose $P_X \neq 0$. Then the fund flows induced by trading X are those in the above case with an unfunded security, plus a "round trip" for P_X , i.e., transferring P_X from one investor to another at $t = 0$ and then transferring it back at $t = 1$. Hence, the total flows in the case with a funded security is always higher than or equal to that in the case with a non-funded security. Equality occurs when one investor pays the other at $t = 0$, and then happens to lose all his wealth to the other investor at $t = 1$. Moreover, investors may have to borrow to trade a funded security and so induce even more fund flows. The flows induced by trading other securities are not affected by contract X . Therefore, the total fund flows induced by a funded security is always higher than or equal to that induced by an unfunded one.

Proof of Proposition 4

Directly differentiating S leads to all results except those in item 4. To prove item 4, we derive the equilibrium prices when the total supply of asset A is $\beta + \beta^*$. Results in item 4 can be obtained by taking β^* to zero.

Proof of Propositions 5 and 6.

The proof is similar to that of Proposition 1. We first calculate the equilibrium prices based on the portfolio holdings described in items 2 and 3. The market clearing condition in the market for asset V is:

$$x_o^* \frac{e + \beta P_A}{P_V - V_d} = x_p^* \frac{e + \beta P_A}{V_u - P_V}. \quad (27)$$

Similarly, the market clearing condition in the market for asset A is:

$$y_o^* \frac{e + \beta P_A}{P_A - (V_d - \Delta)} + (\alpha_o - x_o^* - y_o^*) \frac{e + \beta P_A}{P_A - L^*} = \beta. \quad (28)$$

Suppose the loan contract in equilibrium is such that the borrower promises to pay back Y^* for the loan (L^*, A) . The market clearing condition for the loan market is

$$(\alpha_o - x_o^* - y_o^*) \frac{e + \beta P_A}{P_V - L^*} L^* = (\alpha_p - x_p^*) (e + \beta P_A). \quad (29)$$

Optimistic investors being indifferent about the three strategies in item 2 of Proposition 5 implies

$$\frac{\mathbf{E}_o \tilde{V} - V_d}{P_V - V_d} = \frac{\mathbf{E}_o \tilde{A} - (V_d - \Delta)}{P_A - (V_d - \Delta)}, \quad (30)$$

$$\frac{\mathbf{E}_o \tilde{V} - V_d}{P_V - V_d} = \frac{\mathbf{E}_o \left[\max(\tilde{A} - Y^*, 0) \right]}{P_A - L^*}. \quad (31)$$

Pessimistic investors being indifferent about the two strategies in item 2 of Proposition 5 implies

$$\frac{V_u - \mathbf{E}_p \tilde{V}}{V_u - P_V} = \frac{\mathbf{E}_p \left[\min(Y^*, \tilde{A}) \right]}{L^*}. \quad (32)$$

Note that from equation (31) we can obtain Y^* as a function of L^* . We denote it as $Y^* = f_1(L^*)$. Investor o is happy to be the borrower of the loan contract (L^*, A) if

$$Y^* \leq f_1(L^*). \quad (33)$$

Similarly, from equation (31), we can obtain Y^* as a function of L^* . We denote it as $Y^* = f_2(L^*)$.

Investor p is happy to be the lender of the loan contract (L^*, A) if

$$Y^* \geq f_2(L^*). \quad (34)$$

One necessary condition for L^* and Y^* to satisfy both (33) and (34) is

$$f_1'(L^*) = f_2'(L^*). \quad (35)$$

Rearranging equations (27)–(32) and equation (35), we obtain the seven equation system:

equations (16)–(17) and the following five

$$x_o^* = \frac{\frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \alpha_p \frac{V_d-\Delta}{L^*}}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)}, \quad (36)$$

$$x_p^* = z^* x_o^*, \quad (37)$$

$$L^* = \frac{1}{1 - h_p} \frac{z^*}{z^* + 1} \mathbf{E}_p \left[\min \left\{ \tilde{A}, Y^* \right\} \right], \quad (38)$$

$$Y^* = V_d + F_D^{-1} \left(\frac{1}{1 - h_p} \frac{1 - h_p - h_o z^*}{1 - h_o - h_o z^*} \right), \quad (39)$$

$$y_o^* = \frac{1}{z^* + 1} \frac{V_u - V_d + \frac{\Delta}{h_o}}{L^* - (V_d - \Delta)} \left[\frac{(1 + z^*) \left(1 - \frac{V_d-\Delta}{L^*}\right)}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)} \alpha_o - \frac{\beta(P_A - L^*)}{e + \beta P_A} - \frac{\frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \frac{V_d-\Delta}{L^*}}{1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right)} \right]. \quad (40)$$

Define α_2 as

$$\alpha_2 \equiv 1 - \frac{\frac{\beta(P_A - L^*)}{e + \beta P_A} \left[1 + z^* \left(1 - \frac{V_d-\Delta}{L^*}\right) \right] + \frac{e+\beta(V_d-\Delta)}{e+\beta P_A} - \frac{V_d-\Delta}{L^*}}{(1 + z^*) \left(1 - \frac{V_d-\Delta}{L^*}\right)}. \quad (41)$$

In case of $\alpha_1 \leq \alpha_p < \alpha_2$, the equation system has a unique solution. The notional interest rate in equilibrium is then

$$r(L^*, A) = \frac{Y^*}{L^*} - 1. \quad (42)$$

The proof of Proposition 6 is analogous to that of Proposition 5. Now the equation system has one less equation since optimistic investors are indifferent about two, rather than three, strategies. Following the same logic, we obtain

$$L^{**} = \frac{1}{1 - h_p} \frac{x_p^{**}}{x_o^{**} + x_p^{**}} \mathbf{E}_p \left[\min \left(\tilde{A}, Y^{**} \right) \right], \quad (43)$$

$$x_o^{**} = \frac{\alpha_o h_o e}{h_o e + \beta \mathbf{E}_o \left[\max \left(\tilde{A} - Y^{**}, 0 \right) \right]}, \quad (44)$$

$$x_p^{**} = \frac{\alpha_p (1 - h_p) e}{(1 - h_p) e + \beta \mathbf{E}_p \left[\min \left(\tilde{A}, Y^{**} \right) \right]}, \quad (45)$$

and Y^{**} is the unique positive solution to

$$\frac{x_p^{**}}{x_o^{**}} = \frac{1 - h_p}{h_o} \frac{1 - (1 - h_o) F(Y^{**} - V_d)}{1 - (1 - h_p) F(Y^{**} - V_d)}. \quad (46)$$

Hence, the notional interest rate in equilibrium is

$$r(L^{**}, A) = \frac{Y^{**}}{L^{**}} - 1. \quad (47)$$

Proof of Propositions 7 and 8

The calculation of the equilibrium in the benchmark economy without the derivative contract V is similar to that on Propositions 1, 5 and 6 and is reported in the online appendix. To prove Proposition 7, it is sufficient to see that in a subset of $(\underline{\alpha}, \alpha_1)$, the price of asset A in this benchmark economy, P_A^B , is given by

$$P_A^B = \frac{e\alpha_o + (\gamma^B + \beta)(V_d - \Delta)}{\gamma^B + \beta\alpha_p}.$$

where

$$\gamma^B = \frac{e + \beta(V_d - \Delta)}{V_u - V_d + 2\Delta},$$

and the expected utility of an optimistic investor, J_o^B , is given by

$$J_o^B = (e + \beta P_A^B) \frac{\mathbf{E}_o[\tilde{A}] - (V_d - \Delta)}{P_A^B - (V_d - \Delta)} = \frac{h_o}{\alpha_o} [e + \beta(V_d - \Delta)] \frac{\gamma^B + \beta}{\gamma^B} \frac{V_u - V_d + \frac{\Delta}{h_o}}{V_u - V_d + 2\Delta}.$$

In the economy with the derivative V the price of asset A , P_A is given by (13), and the expected utility for an optimistic investor, J_o , is given by

$$J_o = (e + \beta P_A) \frac{\mathbf{E}_o[\tilde{V}] - V_d}{P_V - V_d} = \frac{h_o}{\alpha_o} [e + \beta(V_d - \Delta)] \frac{\gamma + \beta}{\gamma}.$$

It is straightforward to see that

$$P_A \gtrless P_A^B \text{ iff } h_o \lesseqgtr \frac{1}{2}.$$

$$J_o \gtrless J_o^B \text{ iff } h_o \gtrless \frac{1}{2}.$$

Similarly, we can calculate the pessimist's welfare in the economy with and without the derivative V , J_p , and W_p^B , and obtain

$$J_p \gtrless J_p^B \text{ iff } h_p \lesseqgtr \frac{1}{2}.$$

That is, the introduction of the derivative V has a mixed impact on the price of asset A and investors' welfare.

References

- Allen, Franklin and Douglas Gale 1994 *Financial Innovation and Risk Sharing* MIT Press
- Arora, Navneet, Priyank Gandhi, and Francis Longstaff, 2010, Counterparty Credit Risk and the Credit Default Swap Market, working paper.
- Basak, Suleyman and Benjamin Croitoru, 2000, Equilibrium mispricing in a capital market with portfolio constraints, *Review of Financial Studies* 13, 715–748.
- Benston George and Clifford Smith, Jr., 1976, A Transactions Cost Approach to the Theory of Financial Intermediation, *Journal of Finance*, 31 215–231.
- Bhamra, Harjoat and Raman Uppal, 2009, The Effect of Introducing a Non-Redundant Derivative on the Volatility of Stock-Market Returns When Agents Differ in Risk Aversion, *Review of Financial Studies* 22, 2303–2330.
- Banerjee, Snehal and Jeremy Graveline, 2011, Trading in Derivatives When the Underlying is Scarce, working paper.
- Biais, Bruno, Jean-Charles Rochet and Paul Woolley, 2010, Innovations, Rents and Risk, Working Paper
- Brunnermeier, Markus and Lasse Pedersen, 2009, Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22, 2201–2238.
- Brunnermeier, Markus and Wei Xiong, 2011, A Welfare Criterion for Models with Heterogeneous Beliefs, working paper.
- Culp, Christopher and Merton Miller, 1995, Metallgesellschaft and the Economics of Synthetic Storage. *Journal of Applied Corporate Finance* 7, 62-76.
- Dang, Tri Vi, Gary Gorton, Bengt Homstrom, Opacity and the Optimality of Debt for Liquidity Provision, working paper
- DeMarzo, Peter and Darrell Duffie, 1999, A Liquidity-Based Model of Security Design, *Econometrica* 67, 65–99.
- Duffie, Darrell and Rohit Rahi, 1995, Financial Market Innovation and Security Design: An Introduction, *Journal of Economic Theory* 65, 1-42
- Dubey, P., J. Geanakoplos, and M. Shubik, 2005 Default and Punishment in General Equilibrium, *Econometrica*, 73(1): 1-37.
- Fleckenstein, Mattias, Francis A. Longstaff, and Hanno Lustig, 2010, Why does the Treasury issue TIPS? The TIPS-Treasury bond puzzle, Working paper, UCLA Anderson School of Management
- Fostel A., Geanakoplos J. 2008. Leverage Cycles and the Anxious Economy, *American Economic Review* 2008, 98:4, 1211-1244.

- Frazzini, Andrea and Lasse Pedersen, 2011, Embedded Leverage, working paper.
- Garleanu, Nicolae and Lasse Pedersen, 2011, Margin-Based Asset Pricing and Deviations from the Law of One Price NBER Working Papers
- Geanakoplos, John. 1997. Promises, Promises. in *The Economy as an Evolving Complex System II*, ed. W. Brian Arthur, Steven Durlauf, and David Lane, 285-300. Reading, MA: Addison-Wesley.
- Geanakoplos, John. 2003. Liquidity, Default, and Crashes: Endogenous Contracts in General Equilibrium. in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Conference, Vol. 2*, 170-185. Econometric Society Monographs.
- Geanakoplos, John, 2009, The Leverage Cycle, NBER Macroeconomics Annual.
- Geczy, Christopher, David Musto, and Adam Reed, 2002, Stocks are Special Too: An Analysis of the Equity Lending Market, *Journal of Financial Economics* 66, 241–269.
- Gennaioli, N., A. Shleifer, and R. Vishny (2010), .Neglected Risks, Financial Innovation, and Financial Fragility, Working Paper.
- Gennaioli, Nicola, Andrei Shleifer, Robert Vishny, 2011, Neglected risks, financial innovation, and financial fragility NBER working paper
- Gorton, Gary and George Pennacchi, 1990, Financial Intermediaries and Liquidity Creation, *Journal of Finance* 45, 49–72.
- Gorton, Gary and Nicholas Souleles, 2006, “Special Purpose Vehicles and Securitization” chapter in *The Risks of Financial Institutions*, edited by Rene Stulz and Mark Carey (University of Chicago Press).
- Gromb, Denis and Dimitri Vayanos, 2002, Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs, *Journal of Financial Economics*, 66, 361–407.
- He, Zhiguo, and Wei Xiong, 2009, Multi-market Delegated Asset Management, working paper.
- He, Zhiguo, and Wei Xiong, 2010, Equilibrium Debt Financing, working paper.
- Jiang, Wenxi and Hongjun Yan, 2012, Levered ETFs, working paper.
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2010, The Aggregate Demand for Treasury Debt, working paper.
- Kiyotaki, Nobuhiro and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105, 211–248.
- Liu, Jun, and Francis Longstaff, 2004, Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities, *Review of Financial Studies* 17, 611-641.
- Merton, Robert, 1989, On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance, *Geneva Papers on Risk and Insurance*, 225-262

- Miller, Merton, 1986, The Behavioral Foundations of Economic Theory, *Journal of Business*, S451-S468
- Mitchell, Mark and Todd Pulvino, 2011, Arbitrage crashes and the speed of capital, *Journal of Financial Economics*, forthcoming.
- Rampini, Adriano and S. Viswanathan, 2010, Collateral, risk management, and the distribution of debt capacity, *Journal of Finance* 65, 2293-2322.
- Ross, Stephen, 1989, Institutional Markets, Financial Marketing, and Financial Innovation, *Journal of Finance*, 44, 541-556
- Ross, Stephen, 2011, The Recovery Theorem, working paper.
- Santos, Tano and Jose Scheinkman, 2011, Competition among Exchanges, *Quarterly Journal of Economics*, 1027-1061
- Simsek, Alp, 2011, Speculation and Risk Sharing with New Financial Assets, NBER working paper
- Tufano, Peter, 2003, Financial innovation, *Handbook of the Economics of Finance*
- White, Lawrence, 2000, Technological Change, Financial Innovation, and Financial Regulation in the US: The Challenges for Public Policy, *Performance of Financial Institutions: Efficiency: Innovation, Regulation*. Patrick Harker and Stavros Zenios, eds, 388-415
- Zapatero, Fernando, 1998, Effects of Financial Innovations on Market Volatility When Beliefs Are Heterogeneous *Journal of Economic Dynamics and Control*, 597-626