Collateral-Motivated Financial Innovation*

Ji Shen
London School of Economics

Hongjun Yan
Yale School of Management

Jinfan Zhang
Yale School of Management

February 14, 2012

*We thank John Geanakoplos, Andrew Metrick, Adriano Rampini and seminar participants at Econometric Society Meetings, University of Minnesota Finance Conference and Yale 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 reducing collateral (or margin) requirements for trading (speculation or hedging). We analyze a model with investors with heterogeneous beliefs. The trading need motivates investors to introduce derivatives, which are endogenously determined in equilibrium. The “optimal” security is the one that isolates the variable with disagreement. It is optimal in the sense that any 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 margining 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 motivated by reducing collateral (or margin) requirements 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. 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. This disagreement motivates investors to trade this asset, and possibly to introduce new derivatives to facilitate their trading. Casual intuition suggests that investors should have incentive to 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 with a frictionless collateralization procedure. In particular, 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.¹ 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 Parato 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.² Moreover, it may be hard or too costly for a broker 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, 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 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.

¹This lack of penalty assumption is perhaps most suitable for thinking about the case of security trading, where many positions are set up for hedging 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.

²One 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)).
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 derivative is designed to have zero initial value. 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 extend 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 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 the liquid TIPS, the haircut is around 3% in Fleckenstein, Longstaff, and Lustig (2010).
contrast, to take an equivalent long position in 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 portion of the cash flow is more volatile, since the derivative saves more collateral in that case. Moreover, when investors’ funding liquidity dries up, the basis increases more for assets with more volatile unwanted 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 junk bonds have higher unwanted risks (i.e., non-default-related risks, such as 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 corporate bonds), or when there is a sudden 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, 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, such as options, futures and swaps, have the feature of allow
investors to take on large positions with very little collateral—the salient feature in our model.
Another example 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.³
This practice essentially carves the unwanted cashflow (in this case, the unwanted cashflow is
the counterparty’s assets other than the collateral) out of the transaction, the main insight
from our analysis. 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. In the context of our analysis, we can view creating an SPV as carving out
the unwanted cashflow (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 over time.
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.⁴ 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 the continuing efforts by regulators and market participants in improving
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.

³This 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 one for example is

⁴For more details see the Customer Portfolio Margin User Guide, available from the website of The Options
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)) in financial innovation. 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 heterogenous beliefs or preferences (Zapatero (1998), Bhamra and Uppal (2009), Simsek (2011), Banerjee and Graveline (2011)). While these studies focus on the impact of innovation on volatility and 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 many contexts, such as macro economy (e.g., 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 a derivative is always in the portfolio that demands less collateral—that’s the point of inventing this derivative!

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 and Section 6 studies the impact of financial innovation on the economy. Some further analysis of the model is presented in Section 5, and Section 7 concludes. All proofs are 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 $\beta$ ($\beta \geq 0$) unit of asset $A$, which is a claim to a random cash flow $\bar{A}$ 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

$$\bar{A} = \bar{V} + \bar{U},$$

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

$$\bar{V} = \begin{cases} V_u & \text{with a probability } h_i, \\ V_d & \text{otherwise,} \end{cases}$$

with $V_u > V_d$ and $h_o > h_p$. We use $\alpha_o$ and $\alpha_p$ to denote the population sizes of optimistic and pessimistic investors, respectively and $\alpha_o + \alpha_p = 1$. 
Without loss of generality, we assume \( \bar{U} \) has a mean of zero. In addition, we have the following simplifying assumptions: First, \( \bar{U} \) 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., \( \bar{A} \) is nonnegative. Third, \( \bar{U} \) is independent of \( \bar{V} \). We use \( F(\cdot) \) to denote the cumulative distribution function of \( \bar{U} \), 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 investors would engage in financial innovation to best facilitate their trading and, consequently, how the financial innovation affects 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

\[
    u_i(c) = \begin{cases} 
    h_i c & \text{if } \bar{V} = V_u, \\
    (1 - h_i) c & \text{if } \bar{V} = V_d. 
    \end{cases}
\]

One can interpret this specification as investor \( o \) having some hedging need at the “up state” \( \bar{V} = V_u \). 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. Hence, 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.\(^5\) 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 compensated further 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 Lehman’s counterparty, this company can immediately seize the collateral in the margin account for that trade. If the collateral value is lower than the amount due to the investor, however, it would be very costly for the investor to get compensated by seizing other assets, because he has to go through the lengthy bankruptcy procedure as an unsecured debt holder. For example, the final settlement plan for Lehman bankruptcy was proved 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 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.\(^6\)

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.

\(^5\)See 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).

\(^6\)All 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.
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 $\beta$ 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. Therefore, the collateral equilibrium is identical to the equilibrium without collateral constraints but investors face a non-negative wealth constraint at $t = 1$.

Without frictions in the collateralization process, investors can simply invent a complete set of Arrow securities in this collateral 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 should have incentive to 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 broker 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, 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 proceeds 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. In practice, securities purchased on margin are using those securities themselves as collateral and the margin loan itself is 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, where investors borrow to purchase an asset and use the asset itself as the collateral, rather than corporations using certain assets as collateral to borrow to finance its investments.

Finally, one might think that long and short positions are treated asymmetrically in our model. This is problematic for derivative contracts, where the distinction between long and short positions is immaterial. However, it is easy to see that this is not the case. Suppose a derivative is a claim to a cash flow $-A$ at $t=1$. An investor is paid when “buying” this asset since it has a negative price. In fact, buying this asset on margin is equivalent to shorting asset $A$ and use cash as collateral. That is, long and short positions are treated symmetrically and simply relabeling long and short positions has no impact on the outcome of our model.

### 3.5 Equilibrium with Individual Security Margin

We define the probability space spanned by $\tilde{V}$ and $\tilde{U}$ as $\mathcal{H} \equiv (\{V_d, V_u\} \times [-\Delta, \Delta], F, P_o, P_p)$, where $\{V_d, V_u\} \times [-\Delta, \Delta]$ is the sample space, $F$ is the sigma-algebra generated by $\{V_d, V_u\} \times [-\Delta, \Delta]$, $P_o$ and $P_p$ are the probability measures for the optimists and pessimists, respectively. Any financial security in this economy can be described as 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 $(C, \theta, L)$, with $\theta \geq 0$, to denote a long position of $\theta$ units of asset $C$, which is financed by the borrowing contract $(L, C)$. That is, the investor buys $\theta$ units of asset $C$ and use the position itself as collateral to borrow $\theta L$ 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 $\theta L$ to purchase an asset that is worth $\theta P_C$, 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, the time $t = 1$ payoff to an investor who lends to finance this long position $(C, \theta, L)$ is

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

If an investor takes a short position in an asset, he 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 buys 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 $V_u - \Delta$. That is, the relevant cash flow in this
transaction is the one in equation (5) rather than the “artificial” promise \( \bar{A} \). 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.\(^7\) If an investor shorts one units of asset \( C \), he needs to post max \( C \) cash collateral, where max \( C \) is the maximum value of the cash flow from asset \( C \) at time \( t = 1 \). Therefore, shorting \( \theta \) units of asset \( C \) needs max \( C - P_C \) capital from the short seller and its payoff at time \( t = 1 \) is

\[
W^-(C, \theta) = \theta(\max \bar{C} - \bar{C}), \quad \text{for } \theta \geq 0.
\]  

(6)

To finance a long position in asset \( C \), an investor can choose any loan contracts \((L, C)\) and may choose multiple types of loan contracts to finance different parts of the position. We use \((C, \theta^+_{i,C}(L), L)\), for \( C = A, K, \theta^+_{i,C}(L) \geq 0 \), and \( L \in [0, P_C] \) to describe these long positions,\(^8\) 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 total 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 a 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 long positions in asset \( C \), which are financed by investor \( i \) and held by other investors; 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 \( L \leq x \) and held by other investors. 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 \( \eta_i \) 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

\[
W_i = \sum_{C = A, K} \left( \int_0^{P_C} W^+(C, \theta^+_{i,C}(L), L) dM^+_{i,C}(L) + W^-(C, \theta^-_{i,C}) + \int_0^{P_C} X(C, \theta^*_{i,C}(L), L) dM^*_{i,C}(L) \right) + \eta_i. \quad (7)
\]

\(^7\)For example, we can redefine the promise of \( \bar{A} \) with \( V_i - \Delta \) cash collateral as the promise of the cash flow in equation (5) with \( V_i - \Delta \) cash collateral.

\(^8\)Since there is no penalty to default in this economy, no investor can borrow more than the collateral value. So we don’t need to consider the case of \( L > P_C \).
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 \(t = 1\):

\[
\max E_i(W_i) \quad (8)
\]

subject to

\[
\sum_{C=A,K} \left( \int_0^{P_C} \theta^+_{i,C}(L)(P_C-L) dM^+_{i,C}(L) + \int_0^{P_C} L \theta^*_{i,C}(L) dM^*_{i,C}(L) + \theta^-_{i,C} \left( \max \tilde{C} - P_C \right) \right) + \eta_i \leq \epsilon + \beta P_A. \quad (9)
\]

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.

**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 \(\beta\) 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.\(^9\)

\(^9\)One 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.
Due to the disagreement on $\bar{V}$, investors would like to speculate on its value at $t = 1$. It is then natural to conjecture that, in equilibrium, investors would adopt a derivative contract, asset $V$, which is a claim to a cash flow $\bar{V}$ 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

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

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

$$
P_A = \mathbb{E}_o[\bar{A}],
P_V = \mathbb{E}_o[\bar{V}].
$$

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 $\alpha_1$ 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 the follows:

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

$$
P_A = \frac{e\alpha_o + (\gamma + \beta) (V_d - \Delta)}{\gamma + \beta \alpha_p}, \quad P_V = \frac{\gamma \alpha_o}{\gamma + \beta \alpha_p} V_u + \frac{(\gamma + \beta) \alpha_p V_d}{\gamma + \beta \alpha_p}.
$$
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 \((V, \frac{\epsilon + \beta P_A}{P_V - V_d}, V_d)\) and \(r(V_d, V) = 0\).

4. Type-p investors short \(\frac{\epsilon + \beta P_A}{P_V - V_u}\) 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 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 \(\alpha \leq \alpha_p < \alpha_1\), 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 \(V\). Therefore, 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 \(V\).\(^{10}\) Each of them holds \(\frac{\epsilon + \beta P_A}{P_V - V_d}\) 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, all of them fins shorting the derivative \(V\) efficient.

These results highlight the main theme of this paper: one important motivation for financial

\(^{10}\)Note 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 \(A\).”
innovation is to introduce securities that need least collateral and so are most efficient for trading. Assets \(A\) and \(V\) give investors the same exposure to \(\tilde{V}\), which investors are interested in betting on. However, derivative \(V\) needs less capital from the investor and so is more efficient in facilitating the bet. If one takes a long position in asset \(A\), he can only borrow \(V_d - \Delta\) against each unit of asset \(A\) as collateral. In comparison, one can borrow \(V_d\) against the collateral of one unit in derivative \(V\).

Note that the cash flow \(\tilde{U}\) is “unwanted”, i.e., investors are not interested trading \(\tilde{U}\). Asset \(A\) offers a way for investors to make bets on \(\tilde{V}\). However, this is not ideal because asset \(A\) is a bundle of \(\tilde{V}\) and \(\tilde{U}\). By taking a position in \(A\), an investor also gets an exposure to \(\tilde{U}\). Although the investor is risk neutral, he still finds this unappealing, because, due to the unwanted risk in \(\tilde{U}\), the investor has to post more collateral when trading it. This 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.

### 3.6 Optimal Financial Innovation

**Definition 2** A financial innovation \(X\) (a claim to a cash flow \(\tilde{X}\) in \(H\)) is optimal if, in the presence of \(X\), one introduced any other derivative contract \(K\) (a claim to a cash flow \(\tilde{K}\) in \(H\)), 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 derivative 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 $\bar{U}$ happens to be $-\Delta$. 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 example 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 $V$ in our model. For example, if asset $X$ is a claim to a cash flow $\bar{X} = a(\bar{V} + b)$, it serves the same economic function as asset $V$. To see this, we note that $a$ can be normalized to 1 by redefine the size of each unit. So, we only need to consider the case $\bar{X} = \bar{V} + b$. The only difference between assets $X$ and $V$ is that asset $X$ pays an extra constant cash flow $b$. Not surprisingly, in 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. The payoff to the position $(X, 1, V_d + b)$ is identical to that to $(V, 1, V_d)$. That is, asset $X$ serves exactly the same economic function as asset $V$ for any value $b$.

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 fund from one account to another. Specifically, the transaction cost for transferring $M$ dollars from one investor to another is $kM$ dollars, where $k$ is positive and close to zero. For simplicity, we assume that investors have already paid a fixed fee to a third party (a dealer) for the fund transferring service before time 0. Therefore, at $t = 0$, investors are facing the same problem as analyzed before. The objective of the dealer is to minimize the transaction cost by “fine-tuning” the derivative contract: When investors are indifferent about a number of derivatives, the dealer can pick the one he prefers. Since investors are facing the same problem, the equilibrium prices are the same as those in Proposition 1. Investors are indifferent about the
above-mentioned derivative $X$ for all $b$. The following proposition shows that a unique derivative contract will be adopted.

**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 the so-called “unfunded” securities, with the name highlighting the fact that investors can establish their positions without paying at the inception and only need to post margin 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 there is a chance for him 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 extend there is an infinitesimal cost of transferring funds, the unfunded security would be strictly preferred.

### 3.8 General Discussions

This collateral view of financial innovation highlights the common theme behind a variety of financial innovations. For example, many successful derivative contracts have the feature of increasing effective leverage. Options are levered positions in the underlying assets. Unfunded securities, such as futures and swaps, fit our model even more closely. These contracts allow investors to take on large positions with very little collateral.

This 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 US: when the counterparty of a derivative or repo transaction defaults, investors can get compensated by simply seizing the collateral posted in the transactions, 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. Essentially, this practice carves the unwanted cash flow (in this case, the unwanted cash flow is the counterparty’s assets other than the collateral) out of the transaction, the main insight analyzed in our model.

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 carving out the unwanted cash flow (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.\textsuperscript{11} 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 the continuing efforts by regulators and market participants in improving 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.

4 Violation of the Law of One Price

Our model also has a number of implications on asset pricing. In particular, as noted in our previous analysis, holding asset $A$ is inefficient due to its higher collateral requirement. In

equilibrium, therefore, to induce an investor to hold asset \( A \), 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 a 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), 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 this bond plus a Treasury bond. The empirical evidence suggests, however, 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 \( \bar{U} \) 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 lots 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. We relegate more detailed discussion on this to the Appendix.

4.1 Model Implications

We use $B$ to denote the basis, the price difference between asset $A$ and its replicating portfolio:

$$B = P_V + P_U - P_A.$$ Naturally, the basis can be decomposed into two components:

$$B = S + P_U,$$ (15)

where $S = P_V - P_A$. The first component $S$ reflects the value of saving collateral. Even though all investors are risk neutral, they still value $A$ less than $V$, since $V$ 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. 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 $\beta^*$ 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: $\frac{\partial^2 S}{\partial \beta^2 \partial \Delta} > 0$;
(5) $S$ increases when there is more trading need in the economy: $\frac{\partial S}{\partial h_0} > 0$, $\frac{\partial S}{\partial \Delta} < 0$.

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 and need more 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., $\Delta$ 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 $\Delta$).

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 measures how much 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 $\alpha_p$. This is because, in the case analyzed above, the investors are predominately optimists. An increase in pessimists’ population size $\alpha_p$ 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 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 bond have a higher unwanted risk, that is, there is more non-default-related risks in junk bonds(e.g., liquidity risk). Moreover, with the financial crisis unfolding, the basis for junk bonds increases more than that for investment grades 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 a 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 is more collateral efficient in the sense that it 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. That is, 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 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 tight.

Proposition 4 also offers a number of testable predictions. Result (2) suggests that other things equal, the corporate-bond-CDS basis should be larger for bonds with higher unwanted risks. This implies, for example, that the basis should be larger for corporate bonds with longer maturities, or when the riskless interest rate volatility is higher. Moreover, Result (4) implies that the corporate-bond-CDS basis should increase when there is a positive supply shock to the corporate bond (e.g., when a large bond holder is forced to liquidate). This supply shock impact should be stronger for bonds with longer maturities, or when the riskless interest rate volatility is larger. Finally, Result (5) implies that, all else being equal, the corporate-bond-CDS basis should be larger when trading motive is stronger. Similarly, for the case of TIPS, our model implies that the price discrepancy between TIPS and inflation swaps increases with interest rate volatility and the maturity of the TIPS (Result (2)). A positive supply shock in TIPS increases the price discrepancy, especially for those with long maturities (Result (4)).

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 other factors that may have contributed 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 protection offered by weakened institutions, this would decrease the inflation swap rates and so decreases the basis, opposite to the observed evidence.

5 Other cases

The analysis so far is focused on the case of $\alpha \leq \alpha_p \leq \alpha_1$, and this section presents the analyze of other cases. Note that in Proposition 1, all the borrowing is riskless. 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 of borrowing more against the collateral. But, in the case $\alpha \leq \alpha_p \leq \alpha_1$, 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 than $V_d - \Delta$ against the collateral of one unit of asset $A$. 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 borrowing backed by asset $A$ has default risk; while in the other case, $\alpha_2 \leq \alpha_p \leq \bar{\alpha}$, all the borrowing backed by asset $A$ has default risk, where the expression for $\alpha_2$ 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 the follows:
1. The prices of assets $A$ and $V$ are given by

$$\begin{align*}
P_A &= \frac{1}{z^* + 1} V_u + \frac{z^*}{z^* + 1} V_d - \left(1 - \frac{1}{h_A} \frac{1}{z^* + 1}\right) \Delta, \quad (16) \\
P_V &= \frac{1}{z^* + 1} V_u + \frac{z^*}{z^* + 1} V_d, \quad (17)
\end{align*}$$

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 \(\frac{z^*(e+\beta P_A)}{x^*_p(P_V-V_d)}\) 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 pessimistic 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 $\alpha_p$. Note that in the equilibrium in Proposition 1, both $P_A$ and $P_V$ decrease in $\alpha_p$. This is intuitive. More pessimists take short positions in $V$, pushing down its price. This attracts more optimists from $A$ to $V$, and hence pushes down the price of $A$ as well. In the case of Proposition 5, however, there is another force. With the increase of $\alpha_p$, more pessimists choose to lend to optimists. This enables optimists to push up $P_A$ and $P_V$. These two forces offset each other in this case, so that $P_A$ and $P_V$ do not depend on $\alpha_p$.

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

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

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

$$P_V = \frac{x_o^{**}}{x_o^{**} + x_p^{**}} V_u + \frac{x_p^{**}}{x_o^{**} + x_p^{**}} V_d,$$  \hspace{1cm} (18)

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 $(A, \frac{e + \beta P_A}{P_V - P_A}, L^{**})$, 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 $\frac{x_p^{**}(e + \beta P_A)}{x_p^{**}(P_V - V_d)}$ 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 $\alpha_p$. At $\alpha_p = \underline{\alpha}$, both $P_A$ and $P_V$ are pinned down by optimists’ expectation $P_A = \mathbb{E}_o[\hat{A}]$ and $P_V = \mathbb{E}_o[\hat{V}]$. 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 $\alpha_p$ changes, as shown in Proposition 5. Finally, in the region $\alpha_2 \leq \alpha_p < \overline{\alpha}$, when $\alpha_p$ 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 loans backed by asset $A$, giving optimists more purchasing power. This second impact dominates for some parameters in the region $\alpha_2 \leq \alpha_p < \overline{\alpha}$, and increases the price of asset $A$.

**Figure 1: Asset Prices and Interest Rates.**

![Asset Prices and Interest Rates](image)

Figure 1: The upper panel plots $P_A$ and $P_V$ against $\alpha_p$, and the lower panel plots the 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 $\hat{U}$ is uniformly distributed.
One can see the above intuition more clearly by examining the interest rates. Across all three cases, \( \alpha \leq \alpha_p \leq \bar{\alpha} \), the loans backed by \( V \) are all riskless and have a zero interest rate. In contrast, the credit risk of the loans backed by \( A \) varies across cases. As shown in the lower panel of Figure 1, the loans backed by asset \( A \) is riskless and have a zero interest rate in the case of \( \alpha \leq \alpha_p < \alpha_1 \). In the case of \( \alpha_1 \leq \alpha_p < \alpha_2 \), however, some of the loans backed by asset \( A \) is riskless and have a zero interest rate while the rest has default risk and has a positive interest rate, which stays a constant throughout the region. Finally, in the case of \( \alpha_2 \leq \alpha_p < \bar{\alpha} \), all loans backed by asset \( A \) have default risk. Note that the interest rate drops when \( \alpha_p \) increases, indicating that when more pessimists compete to lend, they push down the interest rate.

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 optimal financial innovation, i.e., the one with assets \( A \) and \( V \). As a comparison, the second economy does not have the derivative market and is otherwise identical to the first economy. Since the analysis of the economy without \( V \) is similar to that in previous sections, we leave the details of the analysis to the Appendix. In the following, we summarize the impact of financial innovation by comparing across these two economies.

6.1 Asset Price

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

The intuition behind this result 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, and the introduction of the derivative \( V \) can increase, decrease or have no impact on the price.
of asset $A$.

Interestingly, with the presence of asset $V$, the price of asset $A$ can be lower than even the pessimists’ expected value $E_p[A]$. Note that in the economy without asset $V$, the price of asset $A$ is always between $E_p[A]$ and $E_o[A]$, the expected values of the two groups of investors. This is natural. If the price of $A$ were less than $E_p[A]$, for instance, both investors would have incentive to buy it, which would have pushed up the price. In the presence of the derivative $V$, however, Figure 1 shows that when $\alpha_p$ is large, the price of asset $A$ is even lower than the pessimist’s expected value, $E_p[A]$. Although pessimists find it profitable to buy asset $A$, they choose not to do so because they find trading asset $V$ even more profitable.

### 6.2 Welfare

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.\(^\text{12}\) As noted in Section 3.1, one can simply adopt the hedging interpretation of the model. The model under this interpretation is mathematically identical to that under the speculation interpretation. Investors’ welfare under this new interpretation is mathematically identical to their subjective expected utility under the old heterogeneous-belief interpretation.

Intuitively, the introduction of derivative $V$ affect investors’ welfare through two channels. First, the derivative contract $V$ helps investors to transfer their wealth to the states they prefer, and so improves their welfare. Second, as seen in Proposition 7, the introduction of $V$ may have a mixed impact on the price of asset $A$, which, in turn, has a mixed impact on investors initial wealth, leading to a mixed impact on investors’ welfare. To see this more clearly, let’s consider the special case in which asset $A$ has a zero net supply, i.e., $\beta = 0$. In this special case, the second channel is shut down since investors’ initial endowment has only cash. Consequently, financial innovation increases all investors’ welfare in the case of $\beta = 0$. The following proposition formalizes the above intuition.

\(^{12}\)Brunnermeier and Xiong (2011) proposes a solution to welfare analysis with heterogeneous beliefs for some cases.
Proposition 8 In the case of $\beta = 0$, the introduction of $V$ weakly increase all investors’ welfare. In the case of $\beta > 0$, the introduction of $V$ has a mixed impact on investors’ welfare.

6.3 Other Impacts

Our analysis suggests that financial innovation allows investors to effectively take on more leverage. On the one hand, this can lead to more speculation and more volatile wealth fluctuations in the economy. 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. The current discussions on the Greek sovereign debt crisis suggests a strong concern about the speculation enabled by financial innovation. For example, the efforts to push for a “voluntary” writedown are often attributed to the concern that a default would trigger large payments from CDS contracts and lead to chaos. This concern seems less consistent with the premise that most of the CDS positions were established for hedging purpose.

Our analysis also reveals a more subtle impact on the economy. 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 this to a separate study.

7 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.

\footnote{Volunteers Wanted, Economist, January 21, 2012.}
that carve out all “unwanted” cash flows. This financial innovation 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. 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 in this case 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 first characterize the equilibrium under this conjecture and then verify it later. 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 \( x_o \frac{e + \beta P_A}{P_V} \). Similarly, pessimists’ aggregate short position in asset \( V \) is \( \alpha_p \frac{e + \beta P_A}{P_V - V_u} \). So the market clearing condition in the market for asset \( V \) is:

\[
x_o \frac{e + \beta P_A}{P_V - V_d} = \alpha_p \frac{e + \beta P_A}{P_V - P_V}.
\]  
(20)

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

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

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

\[
\frac{E_o[\bar{V}] - V_d}{P_V - V_d} = \frac{E_o[\bar{V}] - (V_d - \Delta)}{P_A - (V_d - \Delta)}.
\]  
(22)

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

\[
\frac{V_u - E_p[\bar{V}]}{V_u - P_V}.
\]  
(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 $E_o[V] < P_V < E_o[\tilde{V}]$. 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. So, the only point left to verify (c) is 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 investor $o$ can’t borrow through the riskless technology and has to borrow from another investor. It is easy to verify that it is not optimal to borrow from another type $o$ investor. So, the only thing left to check is whether an investor $o$ 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

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

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

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 \(\alpha^*\), \(0 < \alpha^* < 1\), such that if \(\alpha_p = \alpha^*\) there is a unique solution for this equation system. We define

\[
\alpha_1 \equiv \alpha^*.
\]  

(26)

The appendix also shows that if \(\alpha < \alpha_2\), 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**

The proof can be constructed in the following 4 steps.

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 2 or more 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 those investors have disagreement. In other words, before \(K\) is introduced into the
economy, some of those investors’ expected utility is $J_1$ and others’ expected utility is $J_2$. This implies 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, it would generate trades. This contradicts the conclusion in Step 3.

**Proof of Proposition 3**

Note that all investors are indifferent about contracts with any value for $b$. So we just need to prove that the bank strictly prefers an unfunded security. 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 bank 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. So the bank strictly prefers the unfunded security.

**Proof of Proposition 4**

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

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

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

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

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

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

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

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

$$\frac{E_o \tilde{V} - V_d}{P_V - V_d} = \frac{E_o \max (\tilde{A} - Y^*, 0)}{P_A - L^*}. \quad (31)$$

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

$$\frac{V_u - E_p \tilde{V}}{V_u - P_V} = \frac{E_p \min (Y^*, \tilde{A})}{L^*}. \quad (32)$$

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

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

Rearranging equations (27)–(32) and equation (35), we obtain the seven equation system: equations (16)–(17) and the following five

$$x_o^* = \frac{e + \beta(V_d - \Delta)}{e + \beta P_A} - \frac{\alpha_p V_d - \Delta}{L^*};$$  \hfill (36)

$$x_p^* = z^* x_o^*;$$  \hfill (37)

$$L^* = \frac{1}{1 - h_p} z^* \left[ \frac{1}{1 - h_p} \frac{1}{1 - h_o} \frac{e + \beta V_d - \Delta}{L^*} \right];$$  \hfill (38)

$$Y^* = V_d + F^{-1} \left[ \frac{1}{1 - h_p} \frac{1}{1 - h_o} \frac{1}{1 - h_o} \frac{e + \beta V_d - \Delta}{L^*} \right];$$  \hfill (39)

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

Define $\alpha_2$ as

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

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

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

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

$$L^{**} = \frac{1}{1 - h_p} \frac{x_p^{**}}{x_o^{**}} E_p \left[ \min (\tilde{A}, Y^{**}) \right];$$  \hfill (43)

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

$$x_p^{**} = \frac{\alpha_p (1 - h_p) e}{(1 - h_p) e + \beta E_p \left[ \min (\tilde{A}, Y^{**}) \right]}.$$  \hfill (45)
and $Y^{**}$ is the unique positive solution to

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

(46)

Hence, the notional interest rate in equilibrium is

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

(47)

**Proof of Propositions 7**

As a benchmark, we calculate the equilibrium in an economy without any financial innovation (details to be added.) The introduction of the derivative $V$ increases (decreases) the price of asset $A$ if $h_o < 1/2$ ($h_o > 1/2$).

**Proof of Propositions 8**

For the case of $\beta = 0$, it is clear that investors’ expected utility is simply their initial cash holding. Once $V$ is introduced, investors’ expected utility cannot go down since they can always choose not to trade. It is straightforward to verify that financial innovation has mixed impacts on welfare if $\beta > 0$. 
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.


Banerjee, Snehal and Jeremy Graveline, 2011, Trading in Derivatives When the Underlying is Scarce, working paper.


Brunnermeier, Markus and Wei Xiong, 2011, A Welfare Criterion for Models with Heterogeneous Beliefs, working paper.


Dang, Tri Vi, Gary Gorton, Bengt Homstrom, Opacity and the Optimality of Debt for Liquidity Provision, working paper


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


Gennaioli, Nicola, Andrei Shleifer, Robert Vishny, 2011, Neglected risks, financial innovation, and financial fragility NBER working paper


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.


Santos, Tano and Jose Scheinkman, 2011, Competition among Exchanges, Quarterly Journal of Economics, 1027-1061


Tufano, Peter, 2003, Financial innovation, Handbook of the Economics of Finance
