FROM CASH-IN-THE-MARKET PRICING TO FINANCIAL FRAGILITY

Franklin Allen
University of Pennsylvania

Douglas Gale
New York University

Abstract
We review some recent research that explores the relationship between asset-price volatility and financial fragility when markets and contracts are incomplete. (JEL: E5, G2)

The cause of a financial crisis "... may be trivial, a bankruptcy, a suicide, a flight, a revelation, a refusal of credit to some borrower, some change of view that leads a significant actor to unload. Prices fall. Expectations are reversed. The movement picks up speed. To the extent that speculators are leveraged with borrowed money, the decline in price leads to further calls on them for margin or cash, and to further liquidation. As prices fall further, bank loans turn sour, and one or more mercantile houses, banks, discount houses, or brokerages fail. The credit system itself appears shaky and the race for liquidity is on." Kindleberger (1978, pp. 107–108).

1. Introduction

In a series of papers published over the last ten years (Allen and Gale 1994, 1998, 2000a, 2000b, 2004, 2005; Gale, 2003, 2004) and a recent book (Allen and Gale 2000c), we described a liquidity-based approach to understanding financial crises. The central idea is that, when markets are incomplete, financial institutions are forced to sell assets in order to obtain liquidity. Because the supply of and demand for liquidity are likely to be inelastic in the short-run, a small degree of aggregate uncertainty can cause large fluctuations in asset prices. Holding liquidity involves an opportunity cost and the suppliers of liquidity can only recoup this cost by buying assets at fire-sale prices in some states of the world; so, the private provision of liquidity by arbitrageurs will always be inadequate to ensure

complete asset-price stability. As a result, small shocks can cause significant asset-price volatility. If the asset-price volatility is severe enough, banks may find it impossible to meet their fixed commitments and a full-blown crisis will occur.

To illustrate these ideas, we use a model with four essential elements: (i) There is a trade-off between asset returns and asset maturity: short-term assets mature quickly but have low returns; long-term assets have higher returns but take longer to mature. (ii) Following Diamond and Dybvig (1983), we model consumers' liquidity preference as uncertainty about time preference. Ex ante, consumers are identical; ex post, they are either early consumers, who only value immediate consumption, or late consumers, who only value future consumption. (iii) Intermediaries are modeled as risk-sharing institutions that provide liquidity insurance to consumers. Intermediaries pool the consumers' endowments and invest them in a mixture of short-term and long-term assets. They offer consumers risk-sharing contracts that provide a better mix of liquidity and returns than consumers could achieve on their own. (iv) Interbank markets allow intermediaries to trade aggregate risks and, in particular, to hedge against unexpected liquidity shocks. Because of the transaction costs of participation, consumers do not participate in the interbank markets.

In Allen and Gale (2004) we investigate the sufficient conditions for efficient risk sharing. When (a) markets for aggregate risk are complete, (b) participation is incomplete, and (c) contracts are complete, a laissez-faire equilibrium is incentive-efficient. The risk-sharing contracts intermediaries offer consumers may be incomplete: for example, a demand deposit offers consumers a fixed amount of money independently of the aggregate state of nature. Nonetheless, incompleteness of contracts does not lead to market failure. When (a) markets for aggregate risk are complete, (b) participation is incomplete, and (c) contracts are incomplete, a laissez-faire equilibrium is constrained-efficient.

Complete markets also have important implications for asset pricing. When markets are complete, intermediaries trade contingent securities to provide liquidity in each state. There is no need to sell assets to obtain liquidity and asset pricing is independent of liquidity needs. When markets are incomplete, however, intermediaries are forced to sell assets in order to obtain liquidity. An increase in liquidity demand increases the quantity of assets supplied to the market, which reduces asset prices. The fall in asset prices may necessitate the supply of an even greater quantity of assets. This "backward bending supply curve of assets" lies behind the phenomenon of financial fragility.

In Allen and Gale (2005) we investigate the relationship between incomplete markets and asset-price volatility, which provides the key to understanding financial fragility in this model. If an aggregate shock requires several intermediaries to sell assets at the same time, the attempt to obtain liquidity may be self-defeating: as the asset sales push asset prices lower, intermediaries are forced to sell even more assets, which exacerbates the decline in prices. An important component

of this argument is that the supply and demand for liquidity are inelastic in the short-run. In extreme cases, the fall in asset prices may make it impossible for intermediaries to meet their short-term commitments, forcing a default. Because of this "multiplier effect", very small shocks can have large effects on asset prices and financial stability.

Although we have used a stylized model to illustrate these ideas, we believe that the lessons are applicable to many areas of the financial system. Liquidity plays a crucial role in a world in which market and contracts are incomplete. Whenever a firm or financial institution has to make a fixed payment, independently of the state of nature, it runs the risk of having insufficient liquidity. When firms and financial institutions are forced to obtain liquidity by selling assets, as we have seen, the suppliers of liquidity will demand a premium in the form of low asset prices in states where the demand for liquidity is high. This general principle, that the supply of liquidity must always be insufficient to prevent asset price volatility in equilibrium, is not restricted to the banking sector. It applies whenever there are incomplete contracts and incomplete markets.

In Section 2 we describe liquidity preferences and the trade-off between liquidity and asset returns. In Section 3 we explain the role of liquidity in asset-price volatility and the phenomenon of cash-in-the-market pricing. Section 4 introduces intermediation and Section 5 argues that intermediation with non-contingent contracts implies a positive probability of financial crisis. Section 6 explores financial fragility and links theories of financial crisis based on (large, exogenous) real business cycle shocks and theories based on self-fulfilling panics or sunspots.

2. Liquidity

There are two basic elements in our account of liquidity. On the demand side, following Diamond and Dybvig (1983), liquidity preference is represented by uncertainty about future time preferences. On the supply side, there are two assets that exhibit a trade-off between asset returns and liquidity.

There are three dates indexed by $t = 0, 1, 2$, a single all-purpose good at each date, and two assets, a short-term asset and a long-term asset. The short asset is represented by a storage technology: one unit at date t is transformed into one unit of the good at date $t+1$, for $t = 0, 1$. The long asset is represented by a productive investment technology with a two-period lag: one unit of the good at Date 0 is transformed into $R > 1$ units of the good at Date 2. Note the trade-off between asset returns and liquidity: the short asset offers immediate but lower returns; the long asset offers higher but delayed returns. Note also that asset returns are assumed to be certain. This is not necessary for the results, but emphasizes the fact that asset returns are not the source of price volatility.

There is a continuum of identical consumers, each with an endowment of one unit of the good at Date 0 and nothing at Dates 1 and 2. At Date 1, each consumer receives a preference shock. With probability λ, he becomes an early consumer who only values consumption at date 1 and with probability $1 - \lambda$ he becomes a late consumer who only values consumption at Date 2. Expected utility is given by

$$u(c_1, c_2, \lambda) = \lambda U(c_1) + (1 - \lambda)U(c_2),$$

where c_1 and c_2 are the consumption of early and late consumers, respectively, and $U(c)$ has the usual neoclassical properties.

The only aggregate uncertainty concerns the demand for liquidity. We assume that λ is a random variable. For simplicity, suppose λ takes two values $0 < \lambda_L < \lambda_H < 1$ with equal probabilities. At Date 0, consumers know the model and the prior distribution of λ. At Date 1, they observe the true value of λ and the individual agents discover whether they are early or late consumers.

If markets were complete, optimal risk sharing could be achieved. Here we assume that markets are incomplete. In fact, we assume that there are only spot markets for goods and assets.

3. Cash-in-the-Market Pricing

The relationship between liquidity and asset prices plays a crucial role in our account of financial fragility. A simple model first introduced in Allen and Gale (1994) illustrates how asset prices depend on the liquidity of the market participants' portfolios, as well as on the traditional factors of productivity and thrift. We referred to this as "cash-in-the-market pricing."

For the purpose of this illustration, we assume that there is no intermediation. At Date 0, consumers invest their endowments in a portfolio consisting of y units of the short asset and $1 - y$ units of the long asset. At Date 1, a consumer discovers whether he is an early or late consumer. If he is an early consumer, he liquidates his portfolio and consumes the proceeds immediately. If he is a late consumer, after possibly rebalancing his portfolio at Date 1, he waits until Date 2 and then liquidates his portfolio and consumes the liquidated value.

Let P_s denote the price of the long asset at Date 1 in state $s = H, L$. Market clearing at Date 1 requires that late consumers are willing to hold the long asset. This in turn implies that

$$P_s \leq R.$$

Otherwise, the return on the short asset is greater than the return on the long asset and no one will be willing to hold the long asset between Dates 1 and 2. For similar reasons,

$$P_s < R$$

implies that the short asset is strictly dominated and no one is willing to hold the short asset between Dates 1 and 2.

Suppose that there is no uncertainty about the asset price at Date 1, that is, $P_H = P_L$. Equilibrium requires the consumers to hold both assets ate Date 0, the short asset because it is necessary to provide consumption at Date 1 and the long asset because it provides the highest return at Date 2. So equilibrium can be achieved at the first date only if the one-period holding returns on both assets are equal, which occurs if and only if

$$P_H = P_L = 1 < R. \tag{1}$$

As we have seen, this implies that the short asset is dominated at Date 1, so no one holds the short asset between Date 1 and Date 2. Consequently, market-clearing in the goods market requires that the supply of the good is equal to the consumption of the early consumers:

$$\lambda_s(y + P_s(1 - y)) = y$$

or

$$P_s = \frac{(1 - \lambda_s)y}{\lambda_s(1 - y)}.$$

But $\lambda_L < \lambda_H$ implies that $P_L > P_H$, contradicting (1). Thus, variations in demand for liquidity must be reflected in asset prices.

In each state s there are only two possibilities: either (i) $P_s = R$ and the demand for consumption is less than or equal to y or (ii) $P_s < R$ and the demand for consumption is equal to y. When $P_s < R$ the asset price is the ratio of the late consumers' supply of "cash" $(1 - \lambda_s)y$ to the early consumers' supply of assets $\lambda_s(1 - y)$. When $P_s = R$, the late consumers holding of "cash" must be greater than or equal to the amount they supply in exchange for assets. Hence, the general formula is given by

$$P_s = \min\left\{R, \frac{(1 - \lambda_s)y}{\lambda_s(1 - y)}\right\}.$$

From this formula, it is clear that $P_H < P_L$ unless $P_H = P_L = R$, as illustrated in Figure 1.

4. Intermediation

The aggregate risk associated with variation in λ_s is nondiversifiable, but the idiosyncractic risk faced by individual consumers is diversifiable. In a given state s, the number of early consumers is known with certainty and the individual is
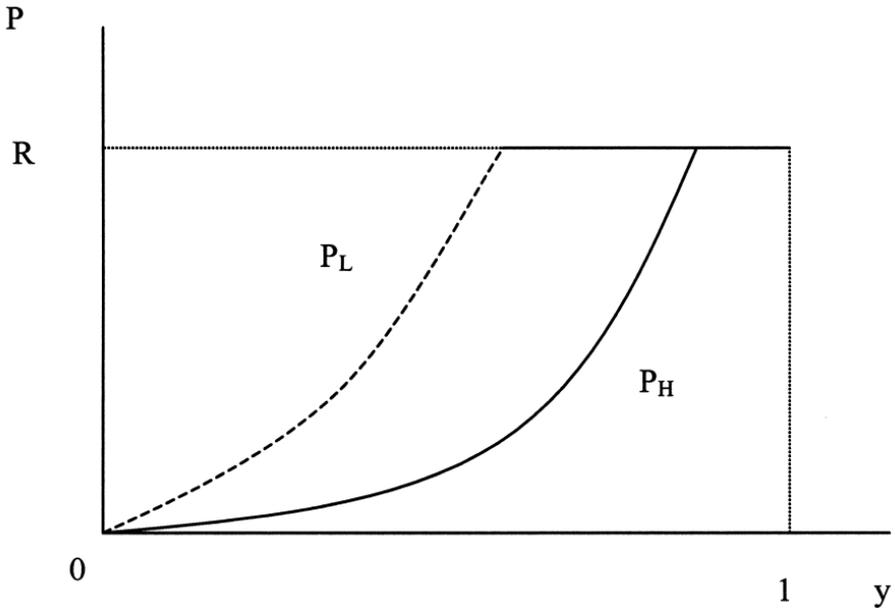
FIGURE 1. Cash in the market pricing.

uncertain whether he will be an early or late consumer. The equilibrium consumption of early and late consumers is given by $c_1 = y + P_s(1 - y)$ and $c_2 = R[(y/P_s) + 1 - y]$, respectively, and $c_1 < c_2$ if $P_s < R$. Risk-averse consumers would be better off ex ante getting the average consumption $y + R(1 - y)$.

As Diamond and Dybvig (1983) showed, intermediaries can provide just this kind of insurance against liquidity shocks by pooling the consumers' resources, investing them in a common portfolio of assets, and offering a better combination of liquidity and returns than individuals could achieve on their own. We assume there is a large number of profit-maximizing intermediaries. Free entry and competition imply that in equilibrium the intermediaries maximize the expected utility of their customers.

At Date 0 consumers deposit their endowments with an intermediary that offers them a deposit contract in exchange. If the intermediary offers them a fully contingent contract, there is no need for or possibility of default. If intermediaries use (noncontingent) deposit contracts, however, the ability of the intermediary to meet its commitments will depend on asset prices and, for some prices, default may be unavoidable.

Suppose the deposit contract promises a fixed amount d if the consumer withdraws at Date 1 and the residual value of the portfolio at Date 2. A late consumer must always receive as much as an early consumer, because he has the option of withdrawing at Date 1 and storing the goods until Date 2. Then the intermediary is solvent at Date 1 if the present value of consumption promised to

consumers does not exceed the present value of assets:

$$\lambda d + \frac{P_s}{R}(1 - \lambda)d \leq y + P_s(1 - y),$$

where P_s/R is the present value of one unit of the good at Date 2. If this inequality cannot be satisfied, the intermediary is insolvent. It must liquidate its entire portfolio and give each depositor $y + P_s(1 - y)$.

Note that the value of the intermediary's portfolio does not change when the intermediary defaults, but the supply of assets to the market does change. If the intermediary is solvent, it supplies an amount of the asset S so that $P_s S + y = \lambda d$. If the Intermediary is insolvent, it supplies $S = 1 - y$. Thus, the supply of assets is

$$S(P) = \begin{cases} \dfrac{\lambda d - y}{P} & \text{if } P \geq \dfrac{y - \lambda d}{(1 - \lambda)d/R - (1 - y)}; \\ 1 - y & \text{if } P < \dfrac{y - \lambda d}{(1 - \lambda)d/R - (1 - y)}. \end{cases}$$

The supply curve is illustrated in Figure 2. This "backward bending supply curve" has three important features: (i) there is a discontinuity at $P = P^*$, where the intermediary is on the brink of insolvency ("at the waterline"); (ii) for values of $P > P^*$, the intermediary is solvent and the supply of the asset is decreasing in
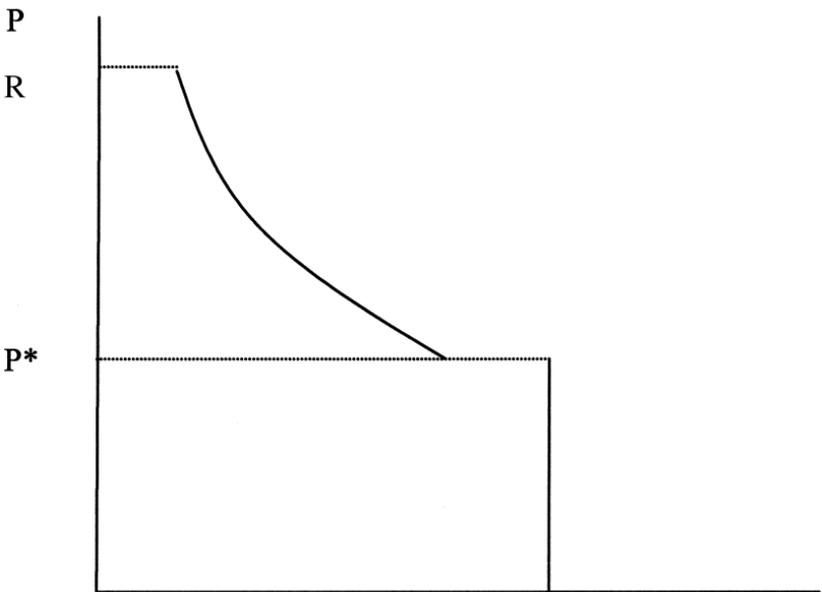

FIGURE 2. The "backward bending supply curve."

P, because the amount of the asset that needs to be sold decreases as the price rises; (iii) for $P < P^*$ the quantity supplied is constant, because the intermediary in default has to sell all of its holding of the long asset.

The critical assumption is that the demand for consumption in Period 1 is greater than the intermediary's holding of the short asset, that is, $\lambda d > y$. The intermediary has to sell off $(\lambda d - y)/P$ units of the long asset to pay for $\lambda d - y$ units of consumption. The lower the asset price P, the more of the asset must be sold. Some of the long asset must be reserved to provide the late consumers with at least $(1 - \lambda)d$ units of consumption. If the asset price P falls far enough, it is impossible to satisfy both early and late consumers. At that point $P = P^*$ and the intermediary is on the verge of default. Any fall in the asset price beyond this point will cause default and the intermediary needs to liquidate its entire stock of the long asset $1 - y$.

5. Crises

What is a crisis? In Allen and Gale (forthcoming), we define it as "a profound fall in asset prices, which affects the solvency of institutions and, in extreme cases, leads to default and collapse." Not all banks default in a crisis, but the fall in asset prices will put them under extreme pressure. Our first result shows that crisis in this sense is unavoidable in our model of intermediation.

Default or Volatility: In equilibrium with aggregate uncertainty about λ, there must be either default or substantial asset-price volatility (or both).

To see this, we assume an equilibrium with no default and show that this implies significant asset-price volatility. Our explanation of asset-price volatility depends crucially on the inelasticity of the demand for and supply of liquidity. The supply of liquidity at Date 1 is perfectly inelastic and equal to y as long as $P_s < R$. Since there is no default, the demand for liquidity is perfectly inelastic and equal to $\lambda_s d$. The market-clearing condition at Date 1 requires

$$\lambda_s d \leq y \tag{2}$$

for $s = H, L$ and $P_s = R$ if the inequality is strict, that is, the asset price must rise until someone is willing to hold the short asset between Date 1 and Date 2. Then the market-clearing condition (2) and the inequality $\lambda_L < \lambda_H$ imply that $\lambda_L d < y$, that is, there is an excess supply of liquidity in state L at any price $P_L < R$. Thus, $P_L = R$. Now, if $P_H \geq 1$ the short asset is dominated at Date 0 and no one will hold the short asset. Thus, market-clearing at Date 0 requires $P_H < 1$. To sum up, the asset-price volatility $P_L - P_H \geq R - 1$ is bounded away from zero as long as there is no default. The difference in prices is equal to the opportunity cost of holding the liquid asset. Note that no assumption has been made about the size of the liquidity shock $\lambda_H - \lambda_L$.

6. Financial Fragility

It is clear that large shocks can have a large impact on asset prices and the possibility of default. The more interesting question is whether the financial system is fragile, in the sense that a small aggregate shock in the demand for liquidity leads to disproportionately large effects in terms of default or asset-price volatility. Allen and Gale (forthcoming) apply the results in the preceding sections to explain financial fragility.

One test for financial fragility is to allow the shocks to become vanishingly small and see whether the effects disappear in the limit. If they do not, we can say that the system is fragile because the shocks are infinitesimal relative to the consequences. We adopt the following stochastic structure for the fraction of early consumers at a single intermediary:

$$\lambda \equiv \alpha + \varepsilon\theta,$$

where $\varepsilon > 0$ is a real number, and α and θ are random variables. The random variable θ represents aggregate uncertainty while α represents idiosyncratic uncertainty. We assume that the idiosyncratic shocks sum to a constant across the economy, so that the average value of α is equal to the expected value with probability one. The parameter ε represents the impact of aggregate uncertainty: as ε converges to zero, aggregate uncertainty disappears.

We have already shown that, in the absence of default, there must be high asset-price volatility, for any value of $\varepsilon > 0$. Thus, we must have either high asset-price volatility or default (or both) when $\varepsilon > 0$. We can use this fact to demonstrate that the asset-price volatility is bounded away from zero even as the amount of aggregate exogenous uncertainty becomes vanishingly small, i.e., as $\varepsilon \to 0$. The proof of this result is by contradiction. If we suppose, contrary to what we want to prove, that asset-price volatility becomes vanishingly small as $\varepsilon \to 0$, then the equilibrium price $P(\theta)$ converges almost surely to a constant. By our earlier result (Default or Volatility), this implies that there must be default in equilibrium for arbitrarily small values of $\varepsilon \to 0$. Allen and Gale (forthcoming) show that the limit of a sequence of equilibria corresponding to values of $\varepsilon \to 0$ is an equilibrium of the limit economy where $\varepsilon = 0$. Thus, default is optimal in the limit. But in the limiting equilibrium, there is no price uncertainty. It can be shown that, as long as the variance of α is not too great, default is never optimal in the absence of price uncertainty. This contradiction forces us to conclude that asset-price uncertainty does not disappear in the limit. This is an example of financial fragility, because a vanishingly small liquidity shock has a large effect on asset prices. We can summarize the discussion as follows:

Financial fragility: Asset-price volatility is bounded away from zero in the limit as $\varepsilon \to 0$. Further, asset-price volatility has real effects as long as α has positive variance.

When $\varepsilon > 0$, variations in θ represent intrinsic uncertainty, that is, uncertainty caused by stochastic fluctuations in the primitives or fundamentals of the economy. When $\varepsilon = 0$, variations in θ represent extrinsic uncertainty, that is, uncertainty that by definition has no effect on the fundamentals of the economy. We call an equilibrium fundamental if the endogenous variables are functions of the exogenous primitives or "fundamentals" of the model (endowments, preferences, technologies). Otherwise, we call the equilibrium a sunspot equilibrium, because endogenous variables may be influenced by extraneous variables (sunspots) that have no direct impact on fundamentals. In the limit economy where $\varepsilon = 0$, an equilibrium is fundamental only if prices do not depend on θ. So unless $P(\theta)$ is almost surely constant, the equilibrium must be a sunspot equilibrium. A crisis cannot occur in a fundamental equilibrium in the absence of exogenous shocks to fundamentals such as asset returns or liquidity demands. In a sunspot equilibrium, by contrast, asset prices fluctuate in the absence of aggregate exogenous shocks and crises appear to occur spontaneously.

Thus there are multiple equilibria in the limit economy with no aggregate exogenous uncertainty. Some of these equilibria are characterized by asset price volatility and financial crises and some are not. Which type of equilibrium is most likely to be observed? To test the robustness of a given equilibrium in the limit economy, we perturb the economy by introducing a small amount of aggregate uncertainty and ask whether there exists an equilibrium of the perturbed economy that is close to the given equilibrium. If the answer is "yes," we describe the equilibrium as robust. Equivalently, we say that an equilibrium is robust if it is the limit of a sequence of equilibria, corresponding to a sequence of perturbed economies, as the shocks become vanishingly small. We already know which equilibria are robust in this sense. We have seen that any equilibrium of the perturbed economy is characterized by asset-price volatility that is bounded away from zero as the aggregate liquidity shocks converge to zero. Thus, a robust equilibrium of the limit economy must have extrinsic uncertainty. Conversely, the fundamental equilibrium of the limit economy is not robust.

Historically, banking panics were often attributed to "mob psychology" or "mass hysteria" (see, e.g., Kindleberger 1978) because there was no obvious economic cause. The modern version of this theory explains banking panics as equilibrium coordination failures (Bryant 1980; Diamond and Dybvig 1983). Out results, which show that some sunspot equilibria are robust and the fundamental equilibrium is not, reinforce the message of the banking panics literature that small shocks can have large consequences. There are important differences, however. Models of banking panics that rely on self-fulfilling prophecies to generate multiple equilibria always have at least one equilibrium which does not involve a panic. Our model, by contrast, makes a firm prediction that every equilibrium must be characterized by extreme asset price volatility as long as there is an arbitrarily small amount of aggregate uncertainty. Moreover, a financial crisis is a systemic phenomenon in our model—a fall in asset prices can occur only if a large number

of banks sell assets, whereas a banking panic may affect a single bank—and from the point of view of the individual bank and its customers, the default is the result of an exogenous change in prices, over which they have no control.

7. Conclusion

We have shown that, where markets and contracts are incomplete and there is no aggregate uncertainty, a "backward bending supply curve" of assets leads to the possibility of multiple equilibria and extrinsic uncertainty. We have also investigated the effect of exogenous aggregate shocks on the demand for and supply of liquidity and shown that even small shocks imply large fluctuations in asset prices because of the inelasticity of demand and supply in the short run. The suppliers of liquidity can only recoup the opportunity cost of holding liquidity by buying assets at fire-sale prices in some states of the world, so the private provision of liquidity by arbitrageurs will always be inadequate to ensure complete asset-price stability. Putting these three components together, we see that whereas price volatility and financial crises can occur in the absence of intrinsic aggregate uncertainty (real exogenous shocks), they must occur if there are arbitrarily small aggregate liquidity shocks. Financial fragility is a robust phenomenon.

We have not explored the reason why markets and contracts are incomplete and why the supply and demand for liquidity are fixed in the short-run. Presumably the reasons for these have to do with microeconomic incentives arising from transaction costs and asymmetric information. Clearly these are important topics for future research, especially if we want to understand the limits of our ability to solve the problems of financial instability through innovation in financial markets and financial contracting. For example, there may be provisions in the bankruptcy code that, while perfectly sensible from the point of view of microeconomic incentives, increase macroeconomic stability.

Understanding the limits of financial markets and financial contracts will also give us a better understanding of the role of the central bank. In particular, it will help us understand whether, in a complex financial system with limited participation, it is sufficient for the central bank to provide adequate liquidity to the financial system as a whole, or whether intervention in specific markets with few informed participants is required.

References

Allen, Franklin and Douglas Gale (1994). "Liquidity Preference, Market Participation and Asset Price Volatility." *American Economic Review*, 84, 933–955.

Allen, Franklin and Douglas Gale (1998). "Optimal Financial Crises." *Journal of Finance*, 53, 1245–1284.

Allen, Franklin and Douglas Gale (2000a). "Financial Contagion," *Journal of Political Economy*, 108, 1–33.

Allen, Franklin and Douglas Gale (2000b). "Optimal Currency Crises." *Carnegie-Rochester Conference Series on Public Policy*, 53, 177–230.

Allen, Franklin and Douglas Gale (2000c). *Comparing Financial Systems.* MIT Press.

Allen, Franklin and Douglas Gale (2004). "Financial Intermediaries and Markets." *Econometrica*, 72, 1023–1061.

Allen, Franklin and Douglas Gale (2005). "Financial Fragility." *Journal of the European Economic Association.*

Bryant, John (1980). "A Model of Reserves, Bank Runs, and Deposit Insurance." *Journal of Banking and Finance*, 4, 335–344.

Diamond, Douglas and Philip Dybvig (1983). "Bank Runs, Liquidity, and Deposit Insurance," *Journal of Political Economy*, 91, 401–419.

Gale, Douglas (2003). "Financial Regulation in a Changing Environment." In *Framing Financial Structure in an Information Environment*, edited by Thomas J. Courchane and Edwin H. Newe. John Deutsch Institute for the Study of Economic Policy, Queen's University.

Gale, Douglas (2004). "Notes on Optimal Capital Regulation." In *The Evolving Financial System and Public Policy.* Bank of Canada.

Kindleberger, Charles (1978). *Manias, Panics, and Crashes: A History of Financial Crises.* Basic Books.