The Economic Journal, **110** (*January*), 236–255. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

BUBBLES AND CRISES*

Franklin Allen and Douglas Gale

In recent financial crises a bubble, in which asset prices rise, is followed by a collapse and widespread default. Bubbles are caused by agency relationships in the banking sector. Investors use money borrowed from banks to invest in risky assets, which are relatively attractive because investors can avoid losses in low payoff states by defaulting on the loan. This risk shifting leads investors to bid up the asset prices. Risk can originate in both the real and financial sectors. Financial fragility occurs when positive credit expansion is insufficient to prevent a crisis.

Financial crises often follow what appear to be bubbles in asset prices. Historic examples of this type of crisis are the Dutch Tulipmania, the South Sea bubble in England, the Mississippi bubble in France and the Great Crash of 1929 in the United States. A more recent example is the dramatic rise in real estate and stock prices that occurred in Japan in the late 1980's and their subsequent collapse in 1990. Norway, Finland and Sweden had similar experiences in the 1980's and early 1990's. In emerging economies financial crises of this type have been particularly prevalent since 1980. Examples include Argentina, Chile, Indonesia, Mexico, and most recently the South East Asian economies of Malaysia, Indonesia, Thailand and South Korea.

These bubbles in asset prices typically have three distinct phases. The first phase starts with financial liberalisation or a conscious decision by the central bank to increase lending or some other similar event. The resulting expansion in credit is accompanied by an increase in the prices for assets such as real estate and stocks. This rise in prices continues for some time, possibly several years, as the bubble inflates. During the second phase the bubble bursts and asset prices collapse, often in a short period of time such as a few days or months, but sometimes over a longer period. The third phase is characterised by the default of many firms and other agents that have borrowed to buy assets at inflated prices. Banking and/or foreign exchange crises may follow this wave of defaults. The difficulties associated with the defaults and banking and foreign exchange crises often cause problems in the real sector of the economy which can last for a number of years.

The Japanese bubble in the real estate and stock markets that occurred in the 1980's and 1990's provides a good example of the phenomenon. Financial liberalisation throughout the 1980's and the desire to support the United States dollar in the latter part of the decade led to an expansion in credit. During most of the 1980's asset prices rose steadily, eventually reaching very

* Helpful comments were received from participants in a CEPR/Bank of England Conference on the Origins and Management of Financial Crises which was held in July 1997, and seminars at the Federal Reserve Bank of New York, the NBER Asset Pricing and Monetary Economics groups, the University of British Columbia and the University of Pennsylvania. We particularly thank Sudipto Bhattacharya, Patrick Kehoe, Frederic Mishkin, Carol Osler, Michael Woodford, Stanley Zin and three anonymous referees for their suggestions. Financial support from the National Science Foundation and the Wharton Financial Institutions Center is gratefully acknowledged.

high levels. For example, the Nikkei 225 index was around 10,000 in 1985. On December 19, 1989 it reached a peak of 38,916. A new Governor of the Bank of Japan, less concerned with supporting the US dollar and more concerned with fighting inflation, tightened monetary policy and this led to a sharp increase in interest rates in early 1990 (see Frankel, 1993; Tschoegl, 1993). The bubble burst. The Nikkei 225 fell sharply during the first part of the year and by October 1, 1990 it had sunk to 20,222. Real estate prices followed a similar pattern. The next few years were marked by defaults and retrenchment in the financial system. The real economy was adversely affected by the aftermath of the bubble and growth rates during the 1990's have mostly been slightly positive or negative, in contrast to most of the post war period when they were much higher.

Many other similar sequences of events can be recounted. As mentioned above, Norway, Finland and Sweden also experienced this type of bubble. Heiskanen (1993) recounts that in Norway lending increased by 40% in 1985 and 1986. Asset prices soared while investment and consumption also increased significantly. The collapse in oil prices helped burst the bubble and caused the most severe banking crisis and recession since the war. In Finland an expansionary budget in 1987 resulted in massive credit expansion. Housing prices rose by a total of 68% in 1987 and 1988. In 1989 the central bank increased interest rates and imposed reserve requirements to moderate credit expansion. In 1990 and 1991 the economic situation was exacerbated by a fall in trade with the Soviet Union. Asset prices collapsed, banks had to be supported by the government and GDP shrank by 7%. In Sweden a steady credit expansion through the late 1980's led to a property boom. In the fall of 1990 credit was tightened and interest rates rose. In 1991 a number of banks had severe difficulties because of lending based on inflated asset values. The government had to intervene and a severe recession followed.

Most other OECD countries experienced similar episodes although they were not as extreme as in Japan and Scandinavia. Higgins and Osler (1997) consider 18 OECD countries and document a significant rise in real estate and stock prices during the period 1984–9. These prices subsequently fell during the period 1989–93. Regression results indicate a 10% increase in real residential real estate prices above the OECD average in 1984–9 was associated with an 8 percent steeper fall than average in 1989–93. Similarly, for equities a 10% increase above the average in the earlier period is associated with a 5% steeper fall in the later period. Higgins and Osler interpret this as suggestive of the existence of bubbles. Investment and real activity were also sharply curtailed during the latter period.

Mexico provides a dramatic illustration of an emerging economy affected by this type of problem. In the early 1990's the banks were privatised and a financial liberalisation occurred. Perhaps most significantly, reserve requirements were eliminated. Mishkin (1997) documents how bank credit to private nonfinancial enterprises went from a level of around 10% of GDP in the late 1980's to 40% of GDP in 1994. The stock market rose significantly during the early 1990's. In 1994 the Colosio assassination and the uprising in Chiapas triggered the collapse of the bubble. The prices of stocks and other assets fell and banking and foreign exchange crises occurred. These were followed by a severe recession.

Kaminsky and Reinhart (1996; 1999) study a wide range of crises in 20 countries including 5 industrial and 15 emerging ones. A common precursor to most of the crises considered was financial liberalisation and significant credit expansion. These were followed by an average rise in the price of stocks of about 40% per year above that occurring in normal times. The prices of real estate and other assets also increased significantly. At some point the bubble bursts and the stock and real estate markets collapse. In many cases banks and other intermediaries were overexposed to the equity and real estate markets and about a year later on average a banking crisis ensues. This is often accompanied by an exchange rate crisis as governments choose between lowering interest rates to ease the banking crisis or raising interest rates to defend the currency. Finally, a significant fall in output occurs and the recession lasts for an average of about a year and a half.

Although the episodes recounted share the same basic progression of the three phases outlined above, they also exhibit differences. One of the most important is the nature of the events associated with the bursting of the bubbles. In many cases the trigger is a change in the real economic environment. An example is the collapse of oil prices in the case of Norway. In other cases the trigger is a result of the expectations about interest rates and the level of credit in the financial system not being fulfilled. An example of this is provided by the collapse of the bubble in Japan in 1990.

How can the basic features of these bubbles and the differing causes of their bursting be understood? There has been a considerable amount of work on bubbles (see Camerer (1989) for an excellent survey) but it can be argued none convincingly capture the sequence of events outlined. Tirole (1982) argued that with finite horizons or a finite number of agents bubbles in which asset prices deviate from fundamentals are not consistent with rational behaviour. The difficulty in reconciling bubbles with rational behaviour resulted in some authors such as De Long *et al.* (1990) developing asset pricing models based on irrational behaviour. Other authors incorporated some form of market imperfection. Tirole (1985), among others, showed that bubbles could exist in infinite horizon models in which all agents are rational. Weil (1987) has shown that bubbles can exist when there is a constant exogenous probability of the bubble collapsing. In his model bubbles crash in finite time with probability one. Santos and Woodford (1997) have argued that the conditions under which bubbles arise in standard general equilibrium frameworks are rather special. These types of models do not provide a good framework for analysing events such as the Japanese, Scandinavian and Mexican bubbles. In the special cases where bubbles do occur, the model does not explain what initiates and ends the bubble.

Allen and Gorton (1993) constructed a model with continuous time and a finite horizon in which an agency problem between investors and portfolio managers could produce bubbles even though all participants were rational.

Allen *et al.* (1993) developed a discrete-time, finite-horizon model where the absence of common knowledge led to bubbles in asset prices. Although these papers show that bubbles can occur because of asymmetric information and agency problems they also fail to capture the typical development of bubbles recounted above. The role of the banking system and the relaxation and tightening of credit is not examined.

The purpose of this paper is to develop a simple formal model in which intermediation by the banking sector leads to an agency problem that results in asset bubbles. Although it has been suggested by Mishkin (1997) and others that problems arising from asymmetric information in the banking system can lead to financial crises, the way in which bubbles arise and their role in crises has not been modelled. There are two main theoretical innovations in the paper:

- The phenomenon of risk shifting or asset substitution is familiar from the corporate finance and credit rationing literature (e.g., Jensen and Meckling, 1976; Stiglitz and Weiss, 1981). However, it has not so far been applied to an asset-pricing context. When investors can borrow in order to invest in pre-existing assets, risk shifting can cause risky assets to be priced above their fundamental value, creating a bubble. This bubble in turn exacerbates the crisis that follows.
- The second innovation is to explore the role of credit expansion in creating bubbles. Credit expansion interacts with risk shifting in two ways. By encouraging investors to fund risky investments at the current date, credit expansion has a contemporaneous effect on asset prices. However, the anticipation of future credit expansion can also increase the current price of assets and it turns out that this may have the greater effect on the likelihood of an eventual crisis.

A number of other authors have stressed the relationship between the banking system and financial crises. McKinnon and Pill (1998) and Krugman (1998) have suggested that it is explicit or implicit government guarantees that lead to risk shifting behaviour and high asset prices. We show that this kind of policy is not necessary for bubbles although it may certainly exacerbate the problem. We suggest that it is uncertainty about the future course of credit creation in the economy and its interaction with the agency problem in intermediation that is crucial for determining the extent of asset price bubbles and ensuing developments. Such uncertainty is often caused by government policies such as financial liberalisation and it is important this is taken into account when such policies are designed.

In this paper we are interested in the first two phases of the process of bubbles followed by a crisis. We do not consider how the banking system deals with sharing the risk associated with crises which is the topic of Allen and Gale (1998). Also, we do not pursue the issue of how default and the resulting disruption in the financial sector spills over into the real economy. There are a number of studies which are complementary to ours. These take as their starting point problems in the financial sector and consider how spillovers to the real sector occur (see, e.g., Bernanke, 1983; Bernanke and Gertler, 1989; Holmstrom and Tirole, 1997).

We start by showing how asset prices are related to the amount of credit and how uncertainty about asset payoffs can lead to bubbles in an intermediated financial system. In this first version of the model default and the resulting financial crisis is caused by low payoffs to the risky assets. In the second part of the paper we develop a dynamic version of the model where it is expectations about the future level of credit that are important in determining asset prices. Here default and crisis result from the actions of the central bank rather than the outcome of any exogenous uncertainty about real economic variables. In the third part, we show that anticipated credit expansion can lead to financial fragility, in the sense that a crisis occurs unless the realised credit expansion is quite large. In other words, a financial contraction is not needed to burst the bubble. The final section of the paper contains a discussion of the results and concluding remarks.

1. Asset Pricing with Uncertainty Generated by the Real Sector

In this section, we analyse a simple model in which the only source of uncertainty is the randomness of real asset returns. Later, the model is extended to allow us to study the dynamic effects of credit expansion.

- There are two dates $t = 1, 2$ and a single consumption good at each date.
- There are two assets, a safe asset in variable supply and a risky asset in fixed supply.
  - *The safe asset*: The safe asset pays a fixed return r to the investor: if x units of the consumption good are invested in the safe asset at date 1 the return is rx units of the consumption good at date 2.
  - *The risky asset*: We can think of the risky asset as real estate or stocks. There is one unit of the risky asset at date 1. If an investor purchases $x \geqslant 0$ units of the risky asset at date 1 he obtains Rx units of the consumption good at date 2, where R is a random variable with a continuous positive density $h(R)$ on the support $[0, R_{MAX}]$ and mean $\overline{R}$.

The safe asset can be interpreted in a number of ways. One possibility is that it is debt issued by the corporate sector. Another possibility is that it is capital goods which are leased to the corporate sector. The investors treat the rate of return as fixed because they are small relative to the size of the corporate sector. In equilibrium, competition will ensure that the rate of return on the bonds or the capital goods leased to the corporate sector is equal to the marginal product of capital.

- The return on the safe asset is determined by the marginal product of capital in the economy. The economy's productive technology is represented by an aggregate production function: $x \geqslant 0$ units of the consumption good at date 1 are transformed into $f(x)$ units of the consumption good at date 2. The production function $f(x)$ is assumed to satisfy the usual neoclassical assumptions, $f'(x) > 0$ and $f''(x) < 0$ for all x, $f'(0) = \infty$ and $f'(\infty) = 0$.

- There is a non-pecuniary cost of investing in the risky asset $c(x)$ which is incurred at the initial date 1. The cost function satisfies the usual neoclassical properties, $c(0) = c'(0) = 0$, $c'(x) > 0$ and $c''(x) > 0$ for all $x > 0$. The risky asset is initially owned by entrepreneurs who supply it inelastically in exchange for the consumption good at date 1.

The purpose of the investment cost $c(x)$ is to restrict the size of individual portfolios and to ensure that, in equilibrium, the borrowers make positive expected profits. There are alternative ways to do this, but this specification leads to a particularly simple analysis.

- There is a continuum of small, risk neutral *investors*. Investors have no wealth of their own, but can borrow from banks to finance investments in the safe and risky assets.
- There is a continuum of small, risk neutral *banks*. The representative bank has $B > 0$ units of the good to lend. Unlike investors, the banks do not know how to invest in the safe and risky assets, that is, they cannot distinguish between valuable and worthless assets. For this reason they have no choice, but to lend to investors.
- The banks and the investors are restricted to using simple debt contracts. In particular, they cannot condition the terms of the loan on the size of the loan or on asset returns.

The assumptions under which it is optimal for banks to write simple debt contracts with investors are well known (see, e.g., Townsend (1979) or Gale and Hellwig (1985)) and we do not discuss them here.

In this paper we have chosen a stark set of assumptions to make the interaction of risk shifting, bubbles and subsequent default as clear as possible. In more realistic models there will be complex interactions and general equilibrium effects, but we are confident that these phenomena will survive in a wide variety of models.

Since there is a continuum of investors and loans cannot be conditioned on their size, investors can borrow as much as they like at the going rate of interest. It follows that in equilibrium the contracted rate of interest on bank loans must be equal to the return on the safe asset. If the rate of interest on bank loans were lower than the return on the safe asset, then the demand for loans by investors would be infinite. On the other hand, if the rate of interest on loans were higher, there would be no investment in the safe asset by investors and so the return on the safe asset would be less than the marginal product of capital since $f'(0) = \infty$. This is inconsistent with our assumption of competition in the corporate sector. Thus in equilibrium the rate of interest on loans must be equal to the return on the safe asset.

Although there is assumed to be a continuum of investors, we shall analyse the behaviour of a representative investor in what follows. This is not just a matter of convenience. It implies that the equilibrium is symmetric and that all investors choose the same portfolio. The fact that all investors are identical *ex post* means that intermediaries cannot discriminate between borrowers by conditioning the terms of the loan on the amount borrowed or any other observable characteristic. X_S and X_R denote the representative investor's holdings of the safe and risky assets, respectively.

Since all investors are treated symmetrically, they will all be charged the same rate of interest r. In principle, it might be possible to condition the interest rate on the amount borrowed, but we shall assume that the exclusive contracts this would require are not feasible. Hence there is linear pricing, that is, the same value of r applies to loans of all sizes. We assume that the banks supply the aggregate amount of loanable funds B inelastically and the rate of interest adjusts to clear the market, that is, to equate the total demand for loans to the amount of (real) credit available.

Because banks use debt contracts and cannot observe the investment decisions of the borrowers, there is a problem of *risk shifting* or *asset substitution.* An investor who has borrowed in order to invest in the risky asset does not bear the full cost of borrowing if the investment turns out badly. When the value of his portfolio is insufficient to repay the bank, he declares bankruptcy and avoids further loss. When the value of his portfolio is high, however, he keeps the remainder of the portfolio's value after repaying the bank. This non-convexity generates a preference for risk.

The optimisation problem faced by the representative investor is to choose the amount of borrowing and its allocation between the two assets to maximise expected profits at date 2. If the representative investor buys X_S units of the safe asset and X_R units of the risky asset, the total amount borrowed is $X_S + PX_R$, where P is the price of the risky asset. The repayment at date 2 will be $r(X_S + PX_R)$. The liquidation value of the portfolio is $rX_S + RX_R$ so the payoff to the investor at date 2 is

$$rX_S + RX_R - r(X_S + PX_R) = RX_R - rPX_R.$$

The optimal amount of the safe asset is indeterminate and drops out of the investor's decision problem, so we can write the investor's problem as follows:

$$\max_{X_R \geqslant 0} \int_{R^*}^{R_{MAX}} (RX_R - rPX_R)\, h(R)\, dR - c(X_R), \tag{1}$$

where $R^* = rP$ is the critical value of the return to the risky asset at which the investor defaults. Because the contracted borrowing rate is equal to the risk-free return, the investor earns no profit on his holding of the safe asset and the default return R^* is independent of the holding of the safe asset.

The market-clearing condition for the risky asset is

$$X_R = 1, \tag{2}$$

since there is precisely one unit of the risky asset. There is no corresponding condition for the safe asset, since the supply of the safe asset is endogenously determined by the investor's decision to invest in capital goods. The market-clearing condition in the loan market is

$$X_S + P = B, \tag{3}$$

since the total amount borrowed is equal to the amount invested in the safe asset X_S plus the market value of the risky asset P. Finally, we have the market-clearing equation for capital goods, which says that the return on the safe asset is the marginal product of capital

$$r = f'(X_S). \tag{4}$$

An *equilibrium* for this model is described by the variables (r, P, X_S, X_R), where the portfolio (X_S, X_R) solves the decision problem (1), given the parameters (r, P), and the market-clearing conditions (2)–(4) are satisfied.

It is straightforward to show that there exists a unique equilibrium (r, P, X_S, X_R) if $\overline{R} > c'(1)$. In this equilibrium the banks supply a fixed amount of credit B inelastically. The contracted interest rate r adjusts to equate the quantity of funds demanded to the quantity supplied. Of course, the realised rate of return will be less than this amount. The typical borrower will default if $R < rP$ so the total return on a loan of one unit is

$$r\Pr(R > rP) + \int_0^{rP} \frac{RX_R + rX_S}{B} h(R)\, dR < r.$$

The 'loss' attributable to the difference between the contracted and the realised rates of return is borne by the banks (or their depositors). It can be thought of as an informational rent that accrues to the investors by virtue of their ability to hide their portfolio from the bank's scrutiny.

The investors' demand for credit is determined by the condition that they make zero profits on the last dollar borrowed. The first-order condition for the maximisation problem (1) equates the expected net return on a unit of the risky asset to the marginal cost of investment, thus ensuring that the zero-profit condition is satisfied for the risky asset. This condition uniquely determines the demand for the risky asset, given the contracted rate of interest r and the asset price P.

Since the rate of return on the safe asset is equal to the contracted rate, the investors make no profits on the safe asset and their demand for the safe asset is indeterminate. The equilibrium amount of the safe asset is determined by the condition that the return on the safe asset is equal to the marginal product of capital, which in turn is a function of the amount of the safe asset.

Substituting from the market-clearing condition $X_R = 1$, the decision problem (1) can be characterised by the first-order condition for the borrower's maximisation problem

$$\int_{R^*}^{R_{MAX}} (R - rP)\, h(R)\, dR = c'(1). \tag{5}$$

Substituting from the budget constraint $X_S = B - PX_R = B - P$, the market-clearing condition for the capital market (4) becomes

$$r = f'(B - P). \tag{6}$$

The two equations (5) and (6) in the two variables (r, P) determine the equilibrium.

The important components of the model, as far as asset pricing is concerned, are the risk shifting problem and the fact that the risky asset is in fixed supply. Borrowers are attracted by the risky asset because they do not bear the loss if they receive a low return, the bank does. On the other hand, when the asset return is high the borrowers receive the surplus and the bank only receives its promised return. This means that the borrowers will bid up the price of the risky asset. As a result, the price of the risky asset is bid up above its 'fundamental' value.

Establishing the fundamental value of the asset is typically a difficult task and depends on the particular circumstances, as Allen *et al.* (1993) have argued. In the current context, where agents are risk neutral, it is natural to define the fundamental as the value that an individual would be willing to pay for one unit of the risky asset if there were no risk shifting, other things being equal. Suppose that a risk neutral individual has wealth B to invest in the safe and risky asset. He would choose a portfolio (X_S, X_R) to solve the problem

$$\begin{array}{ll} \max_{(X_S, X_R) \geqslant 0} & \int_0^{R_{MAX}} (rX_S + RX_R)\, h(R)\, dR - c(X_R) \\ \text{subject to} & X_S + PX_R \leqslant B. \end{array} \tag{7}$$

Comparing the decision problem in (7) to the decision problem in (1) we see that the only differences are that there is no possibility of default in (7). The multiplier on the budget constraint takes the place of the interest rate r in (1). The first-order conditions for this convex problem are necessary and sufficient for a solution:

$$r = \lambda$$

and

$$\int_0^{R_{MAX}} Rh(R)\, dR - rP = c'(X_R). \tag{8}$$

Setting $X_R = 1$ in the first-order condition (8), we can solve it for the fundamental price $\overline{P}$, that is, the price at which an agent who invests his own money would be willing to hold one unit of the risky asset:

$$\overline{P} = \frac{1}{r}[\bar{R} - c'(1)]. \tag{9}$$

Equation (9) defines the fundamental value of the risky asset as the discounted value of net returns. What we would like to show is that the equilibrium price is greater than the fundamental, the classic definition of a bubble.

The equilibrium condition (5) can be rearranged to yield a similar expression:

$$P = \frac{1}{r}\left[\frac{\int_{R^*}^{R_{MAX}} Rh(R)\, dR - c'(1)}{\Pr(R \geqslant R^*)}\right]. \tag{10}$$

Comparing the two pricing kernels (9) and (10), we see that both the numerator and the denominator of (10) are smaller than the corresponding elements of (9). However, the next proposition shows that the two prices can be ranked.

PROPOSITION 1. *There is a bubble in the intermediated equilibrium* (r, P, X_S, X_R). *More precisely, the equilibrium asset price* P *is at least as high as the fundamental price* $\overline{P}$ *and strictly higher as long as the probability of bankruptcy is positive,* $\Pr(R < R^*) > 0$.

Proof. Rewrite (10) as follows:

$$rP = \frac{\int_{R^*}^{R_{MAX}} Rh(R)\,dR - c'(1)}{\Pr(R \geqslant R^*)}$$

$$= \frac{\int_0^{R_{MAX}} Rh(R)\,dR - c'(1) - \int_0^{R^*} Rh(R)\,dR}{\Pr(R \geqslant R^*)}$$

$$= \frac{r\overline{P} - \int_0^{R^*} Rh(R)\,dR}{\Pr(R \geqslant R^*)}. \tag{11}$$

Now

$$\int_0^{R^*} Rh(R)\,dR \leqslant R^* \Pr(R < R^*). \tag{12}$$

Using this together with $R^* = rP$ in (11) gives

$$rP \geqslant \frac{r\overline{P} - rP\Pr(R < R^*)}{\Pr(R \geqslant R^*)}.$$

Since $\Pr(R \geqslant R^*) = 1 - \Pr(R < R^*)$ this simplifies to

$$P \geqslant \overline{P}.$$

If $\Pr(R < R^*) > 0$ then the inequality in (12) is strict and it follows in the same way that $P > \overline{P}$. ■

Proposition 1 shows that the risk shifting that occurs because of the possibility of default leads to prices being higher than the fundamental, which is the discounted value of expected future payoffs.

Because investors are identical everybody will default when $R < R^*$. This widespread default can be interpreted as a financial crisis. Of course in more realistic models with heterogeneous agents only a proportion will default and the proportion defaulting will determine the extent of the crisis.

Proposition 1 illustrates the importance of shocks deriving from the real sector in generating financial crises. For example, Norway's financial problems following the oil price shock can be interpreted as a crisis precipitated by a low realisation of R. What the proposition suggests is that the stage for these problems may have been set when risk shifting led to overinvestment in the risky asset, causing a bubble in asset prices and hence a greater probability of default. The widespread default following the collapse in asset prices caused banks to be insolvent and it was this that led the government to intervene and bail out the banking system.

Because risk shifting behaviour is essential to the creation of a bubble in asset prices, it seems that an increase in the riskiness of the asset returns will increase the size of the bubble. There is a precise sense in which this is true: a mean-preserving spread in the returns to the risky asset increases both the size of the bubble and the probability of default.

To see this, suppose that (r, P) are equilibrium values of the safe return and the price of the risky asset. Now consider a mean preserving spread in the return to the risky asset. There are two cases to be considered. In the first case, the tail of the distribution $h(R)$ on the interval $[0, rP]$ is not affected and in that case there is no change to equilibrium values. In the second case, the lower tail of the distribution on the interval $[0, rP]$ is affected and the entire equilibrium changes as a result. The critical distinction between the two cases is captured by the equilibrium condition

$$\int_{rP}^{R_{MAX}} (R - rP)\, h(R)\, dR = c'(1).$$

The right hand side is a constant, so in equilibrium a mean-preserving spread in $h(R)$ must leave the left hand side constant too. For a fixed value of rP the integral on the left can either increase or stay the same. If it increases, then the value of rP must increase to compensate. Thus, if (r', P') are the equilibrium values corresponding to the new distribution, either $r'P' = rP$ and the equilibrium is essentially unchanged, or $r'P' > rP$.

From the equilibrium condition (6)

$$r = f'(B - P)$$

it is clear that r and P rise and fall together. Therefore, $r'P' > rP$ implies that $r' > r$ and $P' > P$. Furthermore, from the definition of the fundamental value of the risky asset (9)

$$\begin{aligned} \overline{P}' &= \frac{1}{r'}[\overline{R} - c'(1)] \\ &< \frac{1}{r}[\overline{R} - c'(1)] = \overline{P} \end{aligned}$$

because $r' > r$. Thus, the size of the bubble $(P - \overline{P})$ is increased because P rises and $\overline{P}$ falls.

Note also that the probability of default increases. There are two reasons for this. The first is that the riskiness of the risky asset has increased. The second is that rP, the required repayment, has increased. This second, endogenous effect is caused by the risk-shifting behaviour of the investors and amplifies the direct effect of the exogenous increase in risk.

This discussion is summarised in the following proposition.

PROPOSITION 2. *Let* (r, P) *and* (r', P') *denote the equilibrium interest rate and price of risky assets before and after a mean-preserving spread in the distribution of the return* R. *Then either* (a) $(r', P') = (r, P)$ *and the equilibrium is unchanged or* (b) $r' > r$ *and* $P' > P$. *In the latter case, the fundamental value falls* $\overline{P}' < \overline{P}$, *the size of the bubble increases* $P' - \overline{P}' > P - \overline{P}$, *and the probability of default increases.*

2. Asset Pricing with Uncertainty Generated by the Financial Sector

While in some cases it appears that a financial crisis was precipitated by a real shock, in other cases the crisis appears to have been triggered by an event in the financial sector. For example, in many cases financial liberalisation leads to an expansion of credit which feeds a bubble in asset prices. These higher prices are in turn supported by the anticipation of further increases in credit and asset prices. Any faltering of this cumulative process may lead to a crisis. Japan's tightening of credit in 1990 which precipitated the collapse in asset prices provides an example of this.

Of course, if the collapse of asset prices were perfectly foreseen, the bubble would not have been possible in the first place. Backward induction would ensure that the market valuation of assets reflected the fundamental. However, the course of financial liberalisation and credit expansion is never perfectly foreseen. The central bank has limited ability to control the amount of credit. In addition there may be changes of policy preferences, changes of administration, and changes in the external environment, all of which may alter the extent and duration of the credit expansion that feeds the bubble. This uncertainty was particularly great in many emerging economies that underwent financial liberalisation. A similar interpretation may be given to the credit expansions that occurred in developed countries like Japan and that subsequently were summarily cut off.

In the previous section it was shown that there could be a bubble in the sense that the price of the risky asset was higher than its fundamental value. In this section we extend the horizon of the model and show how uncertainty about the extent of credit expansion can increase the magnitude of the bubble.

- There are now three dates $t = 0, 1, 2$ and a single consumption good at each date.

The final two dates 1 and 2 are essentially the same as in the previous model. The main addition is the prior date.

To allow for uncertainty about future credit expansion, we shall assume that B, the amount of credit available for lending to investors, is partially controlled by the central bank. The central bank sets reserve requirements and the quantity of assets available to be used as reserves. By altering one or other of these instruments, the central bank can influence the amount of credit available in the economy. This in turn affects the funds available for investors to purchase the two assets. Because of the uncertainty involved in this process, and this is crucial, investors rationally anticipate an expansion in B but they are uncertain about its exact value. It is now the sequence of credit policies over time, that is, the levels of B_0 and B_1, and the amount of uncertainty associated with B_1 that matters. The following additional assumptions are required.

- At date 0, the level of B_1 is treated by agents as a random variable with a positive, continuous density $k(B)$ on the support $[0, B_{1\,MAX}]$. The price of the risky asset at date 1, $P_1(B_1)$, is therefore also a random variable.
- The safe asset pays $r_t x$ at date $t+1$ if x is invested at date $t = 0, 1$. The owner of the risky asset receives a payoff of $\overline{R}x$ at date 2 if $x \geqslant 0$ is owned at that date.
- There is short term borrowing at dates 0 and 1.
- Entrepreneurs initially own the asset in fixed supply. At date 0 they sell it to investors who hold the representative portfolio from date 0 to date 1. These new investors in turn sell the risky asset at date 1 to the final group of investors who own it until date 2. Investors in the risky asset incur the investment costs $c(x)$ at each date $t = 0, 1$.

As in the two-period model, we can show that the contracted borrowing rate must be equal to the return on the safe asset at each date $t = 0, 1$. Let r_t denote the return on the safe asset at date $t = 0, 1$.

To simplify the analysis and distinguish between the effects of real uncertainty about the asset returns and financial uncertainty about asset prices, we assume that the risky asset has a certain return $\overline{R}$. The risky asset remains risky only because it is a long-lived asset and hence is subject to fluctuations in its price at date 1. Since the safe asset is liquidated after one period, there is no uncertainty about its future value. From the analysis in the preceding section, we know that the equilibrium price of the risky asset at date 1 is given by the formula

$$P_1 = \frac{1}{r_1}[\overline{R} - c'(1)]. \tag{13}$$

Since $r_1 = f'(B_1 - P_1)$ and $pf'(B_1 - p)$ is increasing in p, there is a unique value of P_1 satisfying this equation for each value of B_1. Let $P_1(B_1)$ denote the equilibrium value of the risky asset's price at date 1 when the level of available credit is B_1. Note that $P_1(B_1)$ is continuous and increases without bound if $f'(x) \to 0$ as $x \to \infty$.

Using this expression for the future asset price, we can define an equilibrium at date 0 in the same way as we defined equilibrium in the two-period economy. Consider the representative investor's problem at date 0 (as before, the safe asset drops out of the investor's problem):

$$\max_{X_{0R} \geqslant 0} \int_{B_1^*}^{B_{1\,MAX}} [P_1(B_1) X_{0R} - r_0 P_0 X_{0R}] k(B_1)\, dB_1 - c(X_{0R}) \tag{14}$$

where P_0 is the price of the risky asset at date 0, (X_{0S}, X_{0R}) is the portfolio chosen at date 0, r_0 is the borrowing rate at date 0 and B_1^* denotes the value of B_1 at which the investor is on the verge of default at date 1:

$$P_1(B_1^*) = r_0 P_0. \tag{15}$$

The market-clearing conditions are

$$X_{0R} = 1, \tag{16}$$

$$X_{0S} + P_0 X_{0R} = B_0, \tag{17}$$

and

$$r_0 = f'(X_{0S}). \tag{18}$$

There is no market-clearing condition for the safe asset since its supply is endogenously determined by the investment X_{0S}.

An *equilibrium* is defined by the variables $(r_0, P_0, B_1^*, X_{0S}, X_{0R})$ satisfying (15), the market-clearing conditions (16)–(18) and such that (X_{0S}, X_{0R}) solves the decision problem (14) given the parameters (r_0, P_0, B_1^*).

Again, it is straightforward to show that there exists a unique equilibrium if $\mathrm{E}[P_1(B_1)] > c'(1)$. Making the usual substitutions, we can reduce the set of equilibrium conditions to three:

$$\int_{B_1^*}^{B_{1\,MAX}} [P_1(B_1) - P_1(B_1^*) - c'(1)]\, k(B_1)\, dB_1 = 0 \tag{19}$$

$$r_0 = f'(B - P_0)$$

and

$$P_1(B_1^*) = r_0 P_0.$$

As a benchmark, we first consider an owner investor with B units of wealth and ask at what price $\overline{P}_0$ such an agent would be willing to hold one unit of the risky asset. The decision problem faced by the owner-investor with wealth B_0 is to choose (X_{0S}, X_{0R}) to solve

$$\begin{array}{ll} \max_{(X_{0S},X_{0R})\geqslant 0} & \int_0^{B_{1\,MAX}} [r_0 X_{0S} + P_1(B_1) X_{0R}]\, k(B_1)\, dR - c(X_{0R}) \\ \text{subject to} & X_{0S} + PX_{0R} \leqslant B. \end{array} \tag{20}$$

From the first-order conditions for this problem we can find the fundamental value of the risky asset

$$\overline{P}_0 = \frac{1}{r_0}\{\mathrm{E}[P_1(B_1)] - c'(1)\}. \tag{21}$$

Comparing (21) with the comparable expression for the equilibrium price

$$P_0 = \frac{1}{r_0}\left[\frac{\int_{B_1^*}^{B_{1\,MAX}} P_1(B_1)\, k(B_1)\, dB_1 - c'(1)}{\Pr(B_1 \geqslant B_1^*)}\right] \tag{22}$$

we can prove the following result.

PROPOSITION 3. *Let* $(r_0, P_0, B_1^*, X_{0S}, X_{0R})$ *denote equilibrium values for the intermediated economy and let* $\overline{P}_0$ *be the fundamental price of the risky asset. Then* $P_0 \geqslant \overline{P}_0$ *and the inequality is strict if the probability of bankruptcy* $\Pr(B_1 < B_1^*)$ *is positive.*

Proof. The argument is essentially the same as for Proposition 1. ■

The intuition behind Proposition 3 is the same as for Proposition 1, with uncertainty about B_1 taking the place of uncertainty about R. However, it can be argued the scope for creating bubbles is much greater. The reason is that there is often a great deal of uncertainty about the course of credit expansion and hence how high the bubble may go and when it may collapse. This is particularly true when economies are undergoing financial liberalisation. Thus, the variance of B_1 and $P_1(B_1)$, interpreted as the result of cumulative credit expansion over several years, may be very large. The uncertainty arising from government and central bank policies on credit expansion can dwarf the uncertainty associated with real payoffs on assets. It is the interaction between financial uncertainty and the agency problem in intermediation that leads to the possibility of large deviations of asset prices from fundamentals and subsequent severe financial crises.

3. Financial Fragility

Although Proposition 3 shows how asset prices can become large relative to their fundamentals in intermediated financial systems, it has not yet been shown that credit policies can exacerbate a financial crisis (increase the probability of default). In addressing this issue, the important point is not the existence of the bubble at date 0 but rather the conditions that have to be satisfied at date 1 in order to avoid default by the investors. Even if credit expansion always occurs, that is, $B_1 > B_0$ with probability one, the variability of future credit availability may ensure that a financial crisis occurs. The point is that the expectation of credit expansion is already taken into account in the investors' decisions about how much to borrow and how much to pay for the risky asset. If credit expansion is less than expected, or perhaps simply falls short of the highest anticipated levels, the investors may not be able to repay their loans and a crisis ensues.

To make this more concrete, consider the pricing equation (19) with the substitution of $P_1(B_1^*)$ for $r_0 P_0$ and 1 for X_{0R}.

$$\int_{B_1^*}^{B_{1\,MAX}} [P_1(B_1) - P_1(B_1^*)]\, k(B_1)\, dB_1 = c'(1).$$

As the transactions cost term $c'(1)$ becomes vanishingly small, the left hand side must also vanish, which can only occur if $B_1^* \rightarrow B_{1\,MAX}$. Consequently, there will be a crash unless the expansion of credit is close to the upper bound. The intuitive explanation is that, as transaction costs become less important, competition for the risky asset drives the price up, reducing profits and increasing the incentive for risk shifting.

PROPOSITION 4. *As $c'(1) \to 0$ the default level $B_1^* \to B_{1MAX}$. In other words, credit expansion must not merely be positive but close to the upper bound of the support of B_1 to ensure that a crisis is avoided.*

Note that one does not have to go to this length to produce a high *probability* of crisis. If B_1 has a two-point support concentrated on $\{0, B_{1MAX}\}$ then the probability of a crisis will always be at least $\Pr(B_1 = 0)$, which we can choose as large as we like. The point we want to emphasise is that here we can generate a high probability of crisis *without* resorting to a high probability of significant credit contraction. A crisis can occur even when credit is expanded.

Example: To illustrate the operation of the model consider the following example.

$$B_1 \text{ is uniformly distributed on } [0, 2].$$

$$B_0 = 1;\ f(X_S) = 4X_S^{0.5};\ \overline{R} - c'(1) = 4.$$

Restricting attention to positive prices, it can straightforwardly be shown that

$$P_1(B_1) = 2[(1 + B_1)^{0.5} - 1].$$

By varying the values of $c'(1)$ a number of cases of interest can be generated, as shown in Table 1.

In each of the cases in Table 1, the average level of credit expected by investors next period is the same as the level of credit currently. In the example with $c'(1) = 0.2$ the financial system is robust in the sense that if credit actually remains the same so $B_0 = B_1 = 1 > B_1^* = 0.90$ there will be no default and a financial crisis will be avoided. In the second row where $c'(1) = 0.1$ the financial system is fragile. The amount of credit must be expanded to $B_1 \geqslant B_1^* = 1.21$ if a financial crisis is to be avoided. It is not sufficient to hold it constant or increase it slightly. Finally, in the third case where $c'(1) = 0.01$ the financial system is very fragile. Only a large increase in credit above $B_1^* = 1.74$ will prevent default. Here the probability of a crisis is very high. These examples illustrate that financial crises can occur in a wide variety of circumstances. It is not necessary for there to be a contraction in credit for a

Table 1
A numerical example

$c'(1)$	B_1^*	Probability of a crisis	Intermediated P_0	Fundamental $\overline{P}_0$	Bubble $P_0 - \overline{P}_0$
0.2	0.90	0.45	0.31	0.25	0.06
0.1	1.21	0.61	0.38	0.27	0.11
0.01	1.74	0.87	0.47	0.29	0.18

crisis to be triggered, or even that credit be at or above its expected level, an increase in credit which is simply too small may also result in a crisis.

4. Discussion and Concluding Remarks

The essential feature of the model that explains the existence of bubbles is the risk shifting problem resulting from the inability of lenders to observe how risky borrowers' investments are. The presence of risk shifting causes the price of the asset in fixed supply to be bid up by borrowers. The model presented in this paper has an explicit structure which allows the results concerning bubbles to be illustrated in a straightforward manner. It is meant to be an 'as if' rather than a literal model. The same kind of results will tend to arise whenever there is an agency problem and borrowers have limited liability so that they are only interested in outcomes in the top part of distribution of returns. The particular structure assumed is thus not critical for the results.

Given a certain licence, it is possible to interpret the risky asset in a number of ways. Real estate is an obvious one. Another is stocks. Although in the long run stocks are in variable supply in the short run it takes considerable time to identify profitable opportunities and expand the supply of stocks. Any asset in fixed supply or where the supply is slow to respond to changes in prices will exhibit the kinds of effects discussed.

Our analysis suggests that bubbles will occur when there is considerable uncertainty about real asset payoffs or about credit expansion. We have argued that in particular there is great scope for uncertainty about credit expansion. Financial liberalisation is often a major factor leading to such uncertainty. In designing policies governments and central banks need to take into account the possible impacts of their actions on asset prices if a bubble is to be avoided. It is not simply the level of credit that is important but also the uncertainty of future levels.

One of the simplifications that we have adopted is to focus on the case where all investors use borrowed funds. Our definition of a fundamental considered what the price would be if investors were using their own funds. This raises the question of what would happen if such investors were formally introduced into the model. Depending on the extent of the risk shifting problem and their degree of risk aversion they would either hold less of the risky asset or maybe would even want to short it. If they were risk neutral they would want to short the asset. In order for a bubble to exist in this case there would need to be limitations on short sales or some other limits to arbitrage (see, e.g., Shleifer and Vishny, 1997).

In our model the agency problem arose because of the use of debt contracts and the limited liability these involve. Allen and Gorton (1993) show that a similar agency problem arises when intermediation involves fund managers who bear limited downside risk because the worst that can happen to them is that they are fired. Bubbles can thus arise in a wide variety of financial systems, not just bank-based ones.

As the introduction suggested, bubbles are of interest not only because understanding them requires a theory of asset pricing quite different from the standard ones, but also because in practice they often appear to have significant feedback effects on the real sector of the economy. In order to understand these distortions it is necessary to extend the analysis and model the banking sector more fully.

Banks typically have equity capital and reserves which act as a buffer between the losses incurred on loans and the returns paid to depositors. When returns on the risky asset are low this buffer is depleted and when they are high the buffer is built up to its optimal level. There are many reasons for the existence of these equity buffers, ranging from optimal risk sharing to the avoidance of moral hazard and adverse selection problems between banks and depositors. In addition to the importance of equity capital and reserves there are also default costs which *ex post* at least are borne by the banks. These default costs further deplete bank capital in times of crisis.

The qualitative impact of a financial crisis on the real economy depends on whether the financial crisis is moderate or severe. In a moderate crisis the equity capital and reserves of the banking sector are depleted by default and the deadweight costs of default. In order to restore these to their optimal levels banks may react by reducing their lending. In the case of lending for assets in fixed supply such as real estate and stocks this reduction will help to further lower asset prices. So far, the asset in variable supply has not been associated with any particular sector. A natural interpretation is that it corresponds to the manufacturing sector. In this case the reduction in lending to rebuild bank capital will have a much more severe effect on the economy. If there is a relationship between the funds loaned to manufacturing and employment then the reduction in lending will lead to a recession.

In the case of a severe recession in which many banks fail, losses will be borne by depositors as well as bank shareholders and the stability of the entire banking sector can be threatened. If banks are liquidated, the aggregate capabilities associated with the banks' teams of employees that enables them to distinguish successfully between good assets and bad may be destroyed. In this case total lending may be cut back a very large amount and a severe recession may ensue. Although in recent financial crises, such as those in Scandinavia, governments have prevented the widespread collapse of the financial system by extensive intervention, historically this was not the case. Often banks were allowed to fail in large numbers. In such cases the recessions associated with bubbles were often severe. Recovery is not just a question of rebuilding equity capital and reserves. The banking system itself has to be rebuilt and new teams of employees that can distinguish between good and bad assets have to be developed.

Although a formal model of this relationship between financial and real sectors has not been developed here, Bernanke (1983), Bernanke and Gertler (1989) and Holmstrom and Tirole (1997) among others have discussed such models. Holmstrom and Tirole (1997), for example, develop an incentive model of financial intermediation where intermediaries and firms are credit-constrained. It is shown that the predictions of the model are broadly consistent with interaction between the real and financial sectors in the Scandinavian crises.

In conclusion, this paper has provided a model of bubbles which is consistent with the type of crises observed in Japan, Scandinavia, South East Asia, Mexico and other emerging countries. It was shown how an intermediated financial system could lead to risk shifting and bubbles in asset prices. The relationship between the amount of credit provided by the banking system and the level of asset prices was developed. A fragile regime was identified whereby the central bank must increase the amount of credit by a critical amount in order to avoid a financial crisis, it may not be sufficient simply to increase it.

University of Pennsylvania

New York University

Date of recept of first submission: November 1997
Date of receipt of final typscript: August 1998

References

Allen, F. and Gale, D. (1998). 'Optimal financial crises.' *Journal of Finance*, vol. 53, pp. 1245–84.
Allen, F. and Gorton, G. (1993). 'Churning bubbles.' *Review of Economic Studies*, vol. 60, pp. 813–36.
Allen, F., Morris, S. and Postlewaite, A. (1993). 'Finite bubbles with short sale constraints and asymmetric information.' *Journal of Economic Theory*, vol. 61, pp. 206–29.
Bernanke, B. (1983). 'Non-monetary effects of the financial crisis in the propagation of the great depression.' *American Economic Review*, vol. 73, pp. 257–76.
Bernanke, B. and Gertler, M. (1989). 'Agency costs, collateral, and business fluctuations.' *American Economic Review*, vol. 79, pp. 14–31.
Camerer, C. (1989). 'Bubbles and fads in asset prices: a review of theory and evidence.' *Journal of Economic Surveys*, vol. 3, pp. 3–41.
De Long, B., Shleifer, A., Summers, L. and Waldmann, R. (1990). 'Positive feedback investment strategies and destabilizing rational speculation.' *Journal of Finance*, vol. 45, pp. 379–95.
Frankel, J. (1993). 'The Japanese financial system and the cost of capital.' In (S. Takagi ed.), *Japanese Capital Markets: New Developments in Regulations and Institutions*, pp. 21–77. Oxford: Blackwell.
Gale, D. and Hellwig, M. (1985). 'Incentive-compatible debt contracts: the one-period problem.' *Review of Economic Studies*, vol. 52, pp. 647–63.
Heiskanen, R. (1993). 'The banking crisis in the Nordic countries.' *Kansallis Economic Review*, vol. 2, pp. 13–9.
Higgins, M. and Osler, C. (1997). 'Asset market hangovers and economic growth: the OECD during 1984–93.' *Oxford Review of Economic Policy*, vol. 13, pp. 110–34.
Holmstrom, B. and Tirole, J. (1997). 'Financial intermediation, loanable funds, and the real sector.' *Quarterly Journal of Economics*, vol. 112, pp. 663–91.
Jensen, M. and Meckling, W. (1976). 'Theory of the firm: managerial behaviour, agency cost and ownership structure.' *Journal of Financial Economics*, vol. 3, pp. 305–60.
Kaminsky, G. and Reinhart, C. (1996). 'Banking and balance-of-payments crises: models and evidence.' Working paper, Board of Governors of the Federal Reserve, Washington, D.C.
Kaminsky, G. and Reinhart, C. (1999). 'The twin crises: the causes of banking and balance-of-payments problems.' *American Economic Review*, vol. 89, pp. 473–500.
Krugman, P. (1998). 'Bubble, boom, crash: theoretical notes on Asia's crisis.' Working paper, MIT, Cambridge, Massachussetts.
McKinnon, R. and Pill, H. (1998). 'International overborrowing: a decomposition of credit and currency risks.' Working paper, Stanford University, Palo Alto, California.
Mishkin, F. (1997). 'Understanding financial crises: a developing country perspective.' *Annual World Bank Conference on Development Economics 1996*, pp. 29–61. Washington, D.C.: The International Bank for Reconstruction and Development.
Santos, M. and Woodford, M. (1997). 'Rational asset pricing bubbles.' *Econometrica*, vol. 65, pp. 19–57.

Shleifer, A. and Vishny, R. (1997). 'The limits of arbitrage.' *Journal of Finance*, vol. 52, pp. 35–55.
Stiglitz, J. and Weiss, A. (1981). 'Credit rationing in markets with imperfect information.' *American Economic Review*, vol. 71, pp. 393–410.
Tirole, J. (1982). 'On the possibility of speculation under rational expectations.' *Econometrica*, vol. 50, pp. 1163–81.
Tirole, J. (1985). 'Asset bubbles and overlapping generations.' *Econometrica*, vol. 53, pp. 1499–528.
Townsend, R. (1979). 'Optimal contracts and competitive markets with costly state verification.' *Journal of Economic Theory*, vol. 22, pp. 265–93.
Tschoegl, A. (1993). 'Modeling the behaviour of Japanese stock indices.' In (S. Takagi ed.), *Japanese Capital Markets: New Developments in Regulations and Institutions*, pp. 371–400. Oxford: Blackwell.
Weil, P. (1987). 'Confidence and the real value of money in an overlapping generations economy.' *Quarterly Journal of Economics*, vol. 102, pp. 1–22.