BUBBLES, CRISES, AND HETEROGENEOUS BELIEFS

Wei Xiong

Bubbles, Crises, and Heterogeneous Beliefs
Wei Xiong
NBER Working Paper No. 18905
March 2013
JEL No. G01,G02,G1,G12

<u>ABSTRACT</u>

This paper reviews the quickly growing literature that builds on heterogeneous beliefs, a widely observed attribute of individuals, to explain bubbles, crises, and endogenous risk in financial markets.

Wei Xiong
Princeton University
Department of Economics
Bendheim Center for Finance
Princeton, NJ 08450
and NBER
wxiong@princeton.edu

The history of financial markets has been dotted with episodes of bubbles, during which market values of assets vastly exceeded reasonable assessments of their fundamental value. Asset price bubbles can lead to severe economic consequences ranging from wasteful over-investment and frenzied trading during booms to devastating financial crises and depressed real economies during busts. Economists have emphasized many aspects of bubbles and crises. Minsky (1974) advocated the view that excessive expansion of bank credit due to optimism can fuel a speculative euphoria and slowly lead the economy to a crisis. Kindleberger (1978) stressed that irrationally optimistic expectations frequently emerge among investors in the late stages of major economic booms and lead firm managers to over-invest, over-promise, and over-leverage, which sow the seeds for an eventual collapse after they fail to deliver on their promises. Shiller (2000) highlighted a host of psychological biases people use in forming a feedback mechanism, through which initial price increases caused by certain initial precipitating factors such as new technology innovations feed back into even higher asset prices through increased investor confidence and expectations. Allen and Gale (2007) focused on agency problems of professional managers who actively seek unwarranted risk, which leads to bubbles and crises.

This chapter reviews a quickly growing body of work that was started by Harrison and Kreps (1978) that studies bubbles and crises based on heterogeneous beliefs, a widely observed attribute of individuals. In a market in which agents disagree about an asset's fundamental and short sales are constrained, an asset owner is willing to pay a price higher than his own expectation of the asset's fundamental because he expects to resell the asset to a future optimist at an even higher price. Such speculative behavior leads to a bubble component in asset prices. This approach does not require a substantial amount of aggregate belief distortions to generate a significant price bubble. Instead, the bubble component builds on the fluctuations of investors' heterogeneous beliefs. Even when investors' aggregate beliefs are unbiased, intensive fluctuations of their heterogeneous beliefs can lead to a significant price bubble through frenzied trading (e.g., Scheinkman and Xiong (2003)). This approach is flexible enough to incorporate several important aspects of bubbles and crises, such as over-investment (e.g., Bolton, Scheinkman and Xiong (2006)) and crashes (e.g., Abreu and Brunnermeier (2003) and Hong and Stein (2003)). Heterogeneous beliefs can also lead to credit cycles (e.g., Geanakoplos (2010)). The speculation induced by heterogeneous beliefs also leads to endogenous wealth fluctuations and endogenous risk in general equilibrium settings without short-sales constraints (e.g., Detemple and Murthy (1994) and Kurz (1996)), which can help explain various phenomena such as excess volatility, time-varying risk premium, and high equity premium, which are difficult to explain in standard representative-agent models with a smooth aggregate endowment process. It is also possible to analyze welfare implications of belief distortions based on models with heterogeneous beliefs (e.g., Brunnermeier, Simsek and Xiong (2012)).

This chapter is organized as follows. Section I describes a number of historical bubble episodes with an emphasis on the common characteristics of different episodes. Section II reviews the forces that drive limits of arbitrage, an important ingredient for the rise of price bubbles. This section discusses short-sales constraints observed in many asset markets, non-fundamental risk faced by arbitrageurs in trading against bubbles, and capital constraints that limit the effectiveness of arbitrage trading. Section III discusses different sources of heterogeneous beliefs, including heterogeneous priors, behavioral biases, and frictions in information transmission between investors and advisors. Section IV reviews a number of models that build on heterogeneous beliefs to analyze various aspects of bubbles, such as the rise of bubbles, over-trading, crashes, and over-investment. This section also summarizes several other bubble theories. Section V reviews models of credit cycles based on heterogeneous beliefs. In Section VI, I discuss the large body of literature that explores the general equilibrium implications of heterogeneous beliefs. Section VII discusses a welfare criterion for models with heterogeneous beliefs. Section VIII concludes with some suggestions for future work.

I. Historical Bubbles

Historians of financial markets have vivid accounts of many fascinating bubble episodes across a long time span, across vastly different geographic regions, and across distinct market sectors. In this section, I describe several bubble episodes.

An asset bubble is commonly defined as a period in which an asset's price exceeds its fundamental value. It is difficult to precisely identify a bubble, as any valuation requires a model of fundamentals, which would be subject to concerns about potential model errors. Here, in discussing historical bubbles, I follow Kindleberger's (1978) pragmatic definition of a bubble as "an upward price movement over an extended range that then implodes." Of course, this definition is subject to potential criticism of casual empiricism. As I shall mention later, some of the episodes are more questionable, while the others are more clear-cut.

These episodes serve to highlight several common characteristics of asset bubbles: 1) they tend to coincide with technological or financial innovations; 2) they tend to coincide with frenzied trading between investors, and often active trading of new investors; 3) they are vulnerable to increases in asset supplies; and 4) they may burst suddenly without any warning sign. These common characteristics can guide us in developing a unified framework to analyze bubbles.

The Dutch Tulipmania

The earliest bubble in recorded history was the tulipmania in seventeenth century Holland. Tulips were introduced to Europe in the mid-16th century from the Ottoman Empire. They soon became highly coveted and a status symbol in Holland. The Dutch especially developed a particular taste for exotic bulbs with contrasting colored stripes. According to Mackay (1841), these exotic bulbs triggered speculative frenzies among all classes of the country: "*Nobles, citizen, farmers, mechanics, seamen, footmen, maid-servants, even chimney sweeps and old clothewomen dabbled in tulips.*" The bubble was filled with both technological innovations (i.e., development of new gardening method to grow exotic bulbs) and financial innovations (i.e., the use of new call-option like instruments to facilitate speculation of bulbs). At the peak of the tulipmania, the price of tulip bulbs increased twenty-fold from November 1636 to February 1637, which was followed by even greater declines in the next three months. Garber (2001) examined the evidence and argued that the prices of tulip bulbs were far more rational than was commonly perceived. He pointed out that rare individual bulbs commanded high prices even after the general collapse of bulb prices. But he did not provide a rational explanation for the dramatic twenty-fold increase in tulip-bulb prices in 1637 followed by an even greater decline in prices.

The South Sea Bubble

The South Sea Bubble occurred in Britain in 1720. The South Sea Company was a British joint stock company, which was formed in 1711 to assume the British national debt. As a reward, it was granted a monopoly to trade with Spain's South American colonies. Even though the treaty between Spain and Britain set restrictive quotas on how much the company can trade, the company shrewdly set the stage for a wave of intensive speculation by taking advantage of its cozy relationship with government officials and by exploiting the unlimited potential of the new business opportunity that the trade with the South Seas offered to feed the imagination of its investors. The price of the company's stock rose from £128 in January 1720 to £550 at the end of May 1720. Interestingly, the success of the South Sea Company motivated a large number of other joint-stock companies to float on the stock market. To prevent the large share supply from suppressing its own share prices, the South Sea Company lobbied Parliament to pass the so-called Bubble Act in June 1720, which prohibited unauthorized joint stock ventures. The passing of the Bubble Act further boosted the share price of the South Sea Company to £890 in early June and £1,000 in early August. At that point, the officers and directors of the company started to sell their holdings after realizing the company's prospects could not sustain the market price of the shares. Their selling eventually caused the share price to drop to £100 before the end of the year.

The Roaring Twenties

The Roaring Twenties in the U.S. offered another spectacular bubble in the stock market. This decade was an era of great economic growth propelled by new technologies that made possible mass production of consumer goods such as automobiles and radios. The U.S. emerged as the richest country in the world. However, the Roaring Twenties ended with the Stock Market Crash of 1929 and the onset of the Great Depression. It is difficult to give a precise account of the roles played by fundamentals and speculative mania in driving the stock market boom before the market crash in 1929. Historians (e.g., Galbraith (1997) and White (1990)) summarized evidence along different dimensions that supported the presence of speculative mania as follows: 1) the rapid growth of stock prices near the end of the stock market boom in 1928 and 1929 significantly outpaced the growth of the firms' dividends; 2) many "new-economy" stocks such as RCA, Radio-Keith-Orpheum, and the United Aircraft and Transport Corporation obtained high-flying market valuations despite the dismal hope of any dividends from these stocks in the short- and medium-run; 3) heightened level of trading volume as reflected by a large number of record-setting days on NYSE in 1928 and 1929; 4) the large influx into the market of new investors who had never bought stocks before the boom and whose presence dragged down the overall sophistication of the market participants; and 5) the rapid expansion of brokers' loans which allowed investors to buy stocks on margin.

The Internet Bubble

The biggest bubble of the twentieth century was undoubtedly the Internet bubble in late 1990s. Precipitated by the rapid development of the Internet, a wave of new Internet firms went public. These firms included household names such as Yahoo and Amazon, as well as many other firms that eventually disappeared. The new Internet technology provided these firms with the promise of revolutionizing the way we obtained information, purchased goods and received services. According to Ofek and Richardson (2003), an index tracking the performance of Internet stocks experienced an extraordinary growth of nearly 1000% from October 1998 to February 2000. At the peak in February 2000, the Internet sector had roughly 400 companies and was largely profitless, but commanded an astonishing capitalization of about 6% of the entire U.S. stock market. This sector contributed an even higher fraction---20%---of the publicly traded volume of the U.S. stock market, which directly speaks to the frenzy of trading Internet stocks. The prices of Internet stocks started to decline after February 2000, losing over 80% from the peak by the end of 2000. The losses in market value added up to over $8 trillion, which was larger than half of the one-year output of the U.S. Figure 1, which is taken from Hong and Stein (2007), illustrates the prices and turnover for Internet and non-Internet stocks in 1997-2002.

Figure 1: Prices and Turnover for Internet and Non-Internet Stocks, 1997-2002
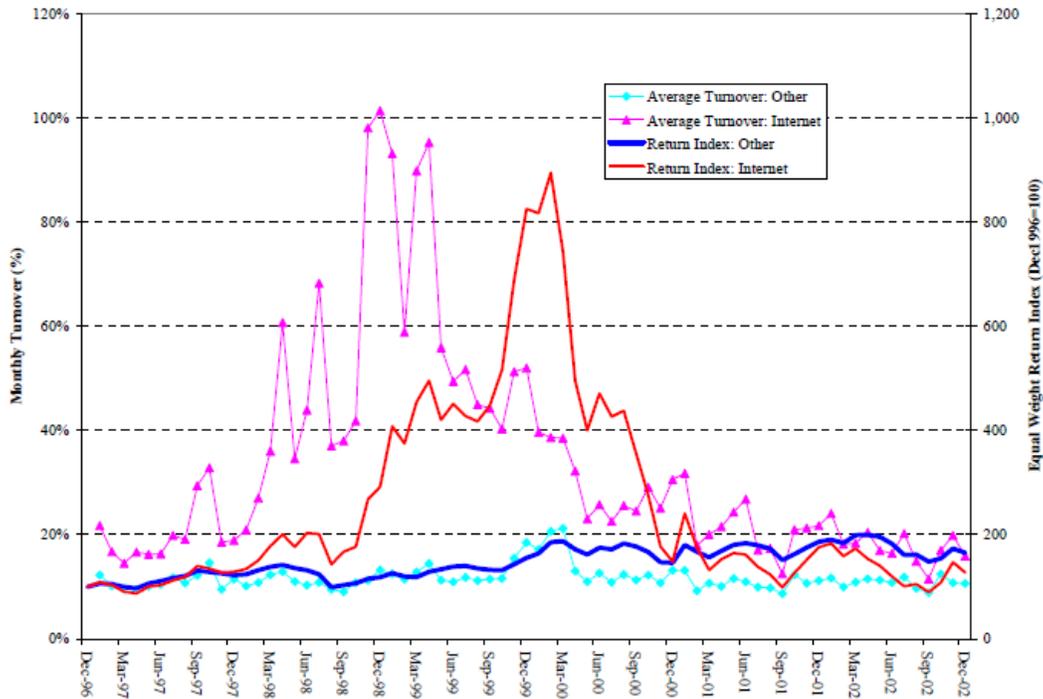

This figure is taken from Figure 1 of Hong and Stein (2007).

Among the many high-flying Internet stocks during this bubble, Palm is particularly illuminating about the irrational exuberance that occurred at the time. Palm was a maker of personal digital assistants, and was initially owned by a company called 3Com. In early 2000, 3Com decided to spin off Palm, among other reasons, to take advantage of the high market valuation of Internet stocks. When it sold 5% of its shares in Palm in a public offering and announced its intention to spin off the remaining shares to the 3Com shareholders, Palm's market valuation was astonishing---the market value of 3Com's holding of the remaining 95% of Palm's shares surpassed that of entire 3Com by over $23 billion! It was as if the rest of 3Com's assets carried a negative valuation of $23 billion. Such outrageous mispricing clearly indicated the presence of a price bubble. Wasn't this a clear arbitrage opportunity? It turned out that shorting Palm stocks, a crucial element in arbitraging the price differential between Palm stocks and 3Com stocks, was not easy. To short Palm stocks, one had to first borrow them in a decentralized search market with a promise to return the borrowed shares to the initial owner later. The small number of Palm's floating shares made it extremely difficult for arbitrageurs to immediately borrow the shares. According to Lamont and Thaler (2003), the short interest eventually went up from less than 20% of the floating shares in March 2000 to 150% in July 2000, which implied that the same share might have been shorted multiple times. During this process, the so-called stub value of 3Com (i.e., the market valuation

5

of 3Com minus its holding of Palm shares) gradually turned from negative $23 billion to positive. See Figure 2 (taken from Lamont and Thaler (2003)) for the plots of 3Com's stub value and Palm's short interest during this period. Lamont and Thaler also documented five other similar carveout cases in which the parent companies carried negative stub values and shorting the subsidiary firms was difficult. In general, during the Internet bubble, shorting Internet stocks was difficult. There were higher shorting interest for Internet stocks, higher borrowing costs for shorting Internet stocks, and greater violation of put call parity for Internet stocks in the options market (e.g., Ofek and Richardson (2003)).

Figure 2: Actual and Synthetic Stub Values of 3Com and Short Interest of Palm in 3/3/2000-7/21/2000
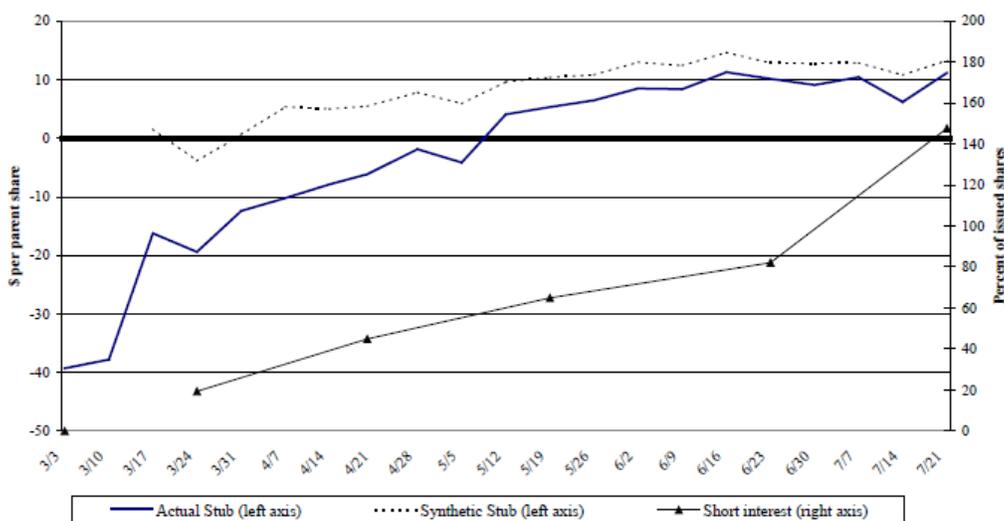
This figure is taken from Figure 5 of Lamont and Thaler (2003).

What caused the Internet bubble to burst? The failure of many Internet firms to produce real earnings after many quarters of financial losses was probably the most important factor that eventually wakened the investors. Continuing losses also exhausted the cash of many Internet firms, which could not sustain their operations without raising additional financing. The hike of the interest rate by the Federal Reverse's monetary policy might have also played a role. The timing of the collapse of Internet stock prices also coincided with the lockup expirations of a large number of Internet firms. Like many previous bubble episodes, the lofty market valuation of Internet stocks instigated numerous public offerings by Internet firms during a concentrated period in 1997-1999. The firm insiders such as entrepreneurs and venture capitalists were initially restricted from selling their shares during the so-called lockup periods. According to Ofek and Richardson (2003), the lockup restrictions of a large number of

6

Internet firms expired in early 2000. The subsequent selling by firm insiders put downward pressure on the share prices and caused the asset float (i.e., the number of tradable shares) of these firms to increase substantially, which in turn made short-selling of their stocks easier.

Taken together, the Internet bubble shared many common characteristics with the previous bubble episodes: highly inflated market prices accompanied by speculative frenzies, short-sales constraints in preventing arbitrage, the lack of favorable economic news, and the largely increased asset float in eventually crashing the prices.

The Chinese Warrants Bubble

Looking back in time, it is difficult to precisely determine whether an asset was in a bubble due to the difficulty in measuring asset fundamentals based on the information investors had at the time of trading. The historical bubble episodes are often defined based on ex post observations of dramatic boom and bust cycles, and thus attract the criticism that one might be able to explain the observed market prices based on particular fundamental valuation methods. For example, Pastor and Veronesi (2006) argue that it is possible to explain the high valuation of Internet firms based on an extended version of the Gordon's growth formula. In the standard Gordon's growth formula, the value of a firm is a convex function of its growth rate, which in turn implies that high uncertainty about the growth rate of Internet firms can help explain their high stock valuation during the Internet bubble.

The recent Chinese warrants bubble provides a unique episode in which one can clearly identify excessive asset valuation relative to asset fundamentals. In 2005-2008, 18 Chinese companies issued put warrants with long maturities ranging from six months to two years. These warrants gave their holders the right to sell the issuing companies' stocks at predetermined strike prices during a pre-specified exercise period. The dramatic boom in the Chinese stock market between 2005 and 2007 pushed most of these put warrants so deep out of the money that they were almost certain to expire worthless. A recent study by Xiong and Yu (2011) documented reliable evidence that the traded warrant prices were substantially higher than warrant fundamentals based on either the standard Black-Scholes option pricing model or other less model-specific upper bounds. Figure 3 (taken from Figure 1 of Xiong and Yu (2011)) illustrates the significantly positive market price of then WuLiang put warrant when its Black-Scholes value dropped to below half of the minimum trading tick of 0.1 penny (one penny is one hundredth of one Yuan, which is less than 0.002 US dollars) for a period of more than 6 months. At its peak, its market price exceeded even its strike price--the highest possible payoff it could generate.

Figure 3: The Price Bubble in WuLiang Put Warrant
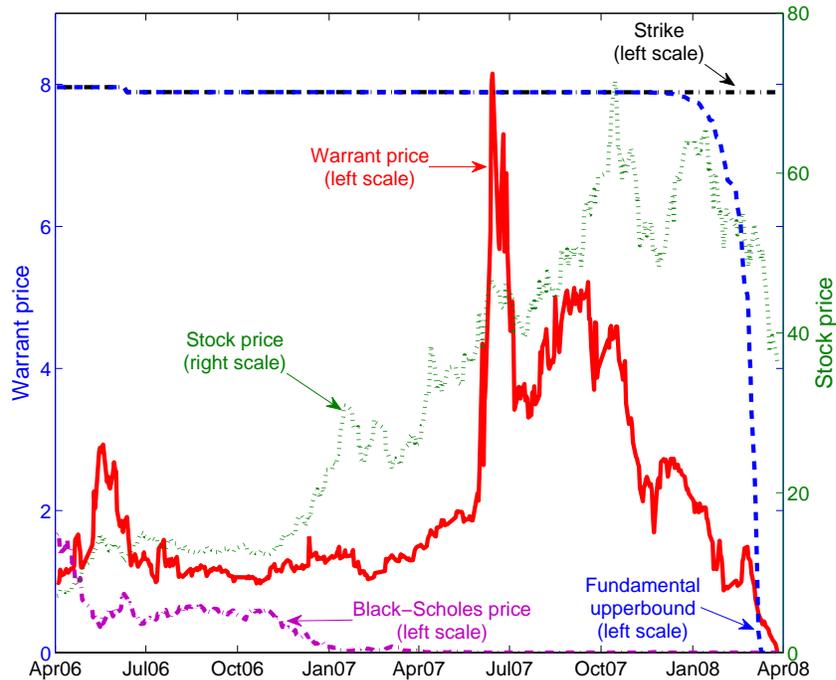
This figure is taken from Figure 1 of Xiong and Yu (2011).

All other warrants, with the exception of one, had a zero-fundamental period in which their Black-Scholes values dropped to below half of the minimum trading tick. This zero fundamental period ranged from 3 to 165 trading days, during which these warrants nevertheless became targets of frenzied speculation. Each warrant had an average daily turnover rate of 328 percent, an average daily volume of 1.29 billion Yuan (roughly 200 million U.S. dollars), and an average return volatility of 271 percent per annum. On an extreme day, the ZhaoHang put warrant had a volume of 45.68 billion Yuan (roughly 7 billion US dollars) even though the warrant was virtually worthless from exercising. On their last trading days, the warrants had an average turnover rate of 1,175 percent in four hours of trading time, which translated into nearly 100 percent turnover every 20 minutes!

This warrants bubble displayed several unique features of bubbles that are observable only in laboratory environments. First, one can reliably measure the warrants' fundamental values to be close to zero by using the underlying stock prices. Second, the publicly observable stock prices also made the warrant fundamentals observable to all market participants. The easily measurable and publicly observable fundamentals made the emergence of the spectacular bubble even more striking. Third, these

8

warrants had predetermined finite maturities. It is commonly perceived that bubbles have explosive price paths before they eventually collapse. In contrast to this common perception, Figure 4 (taken from Figure 3 of Xiong and Yu (2011)) shows that as the number of trading days decreased over time, the average price of the 17 put warrants in their zero-fundamental periods gradually deflated without a sudden crash, while their daily turnover rate and price volatility both increased and, especially, spiked during the last few days.

Figure 4: Warrant Dynamics in the Bubble Sample
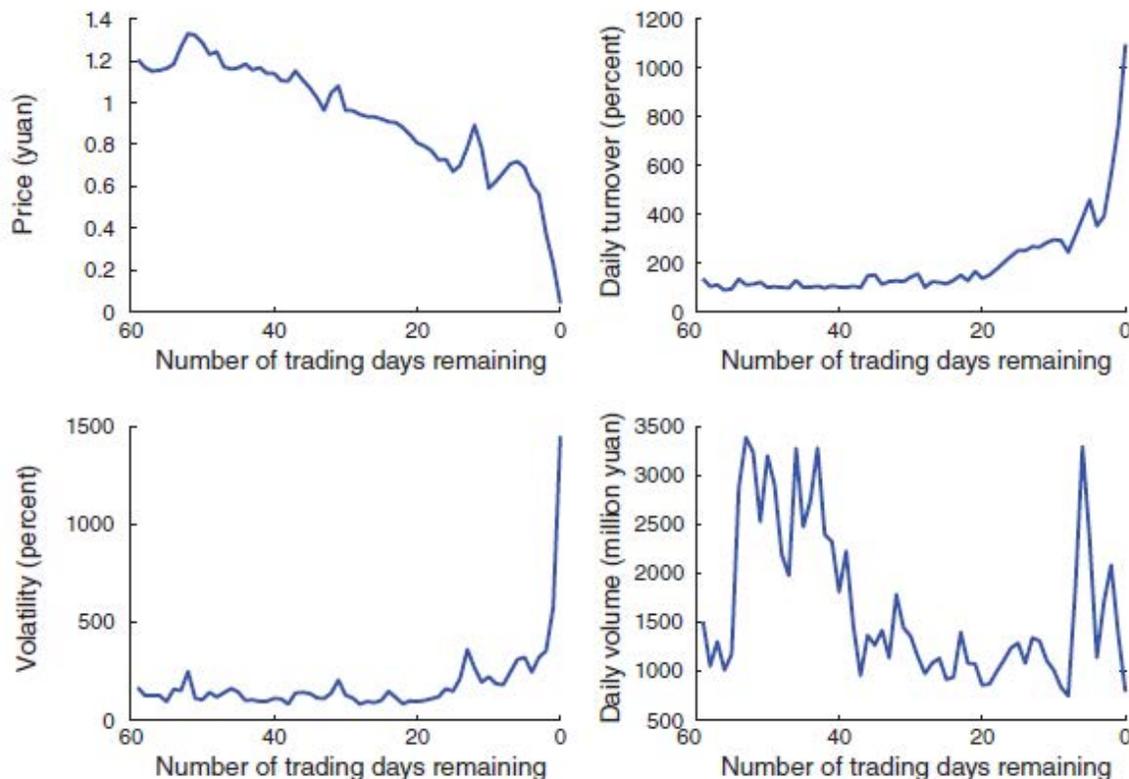
This figure is taken from Figure 3 of Xiong and Yu (2011).

The Chinese warrants bubble reflected characteristics of the previous bubbles in that the extraordinary warrant prices were accompanied by frenzied speculation between investors even to the last trading day and the last trading minute, the rise of the bubble was inherently associated with the restrictive legal ban on short-selling financial securities (including warrants) in China, and the heterogeneity of investors in the newly emerged Chinese financial markets included both sophisticated and totally inexperienced investors.

The U.S. Housing Bubble

The U.S. housing market also experienced a dramatic boom and bust cycle in 2000-2010. Figure 5 (taken from Haughwout et al. (2011)) depicts the average housing prices in the U.S., which had a dramatic appreciation of 100% from 2000 to 2006, with an acceleration of appreciation after 2004, only to be followed by a drop of more than 25% from the peak in 2006 to 2010. This cycle was particularly pronounced in four states---Arizona, California, Florida, and Nevada (the "bubble" states). The prices in these states had an average appreciation of 150% from 2000 to 2006 followed by a drop of nearly 50% from 2006 to 2010. Following Kindelberger's definition of bubbles, I call this dramatic boom and bust cycle a housing bubble.

Figure 5: Housing Prices in US and Four Bubble States (AZ, CA, FL, and NV)
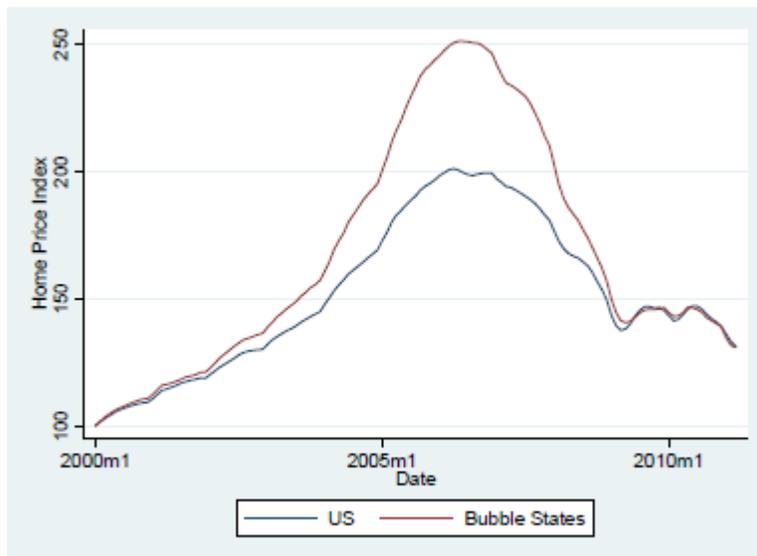
This figure is taken from Figure 1 of Haughwout et al. (2011).

It is widely recognized that this housing bubble lay at the heart of the recent Great Recession. There is ample evidence linking the rise of the housing bubble to the credit market boom caused by the expansionary U.S. monetary policy and lax bank lending standards in early 2000s (e.g., Rajan (2010) and Mian and Sufi (2009)). There is also growing evidence indicting active involvement of "buy and flip" investors during the housing bubble. Haughwout et al. (2011) measured buy and flip investors by new purchase mortgages by borrowers with two, three, and four or more first-lien mortgages on their credit reports. They documented that the fraction of investors in all new purchase mortgages increased from

around 20% in 2000 to a peak of nearly 35% in 2006. The increase in investor share was especially pronounced in the four bubble states, rising from almost 25% in 2000 to 45% in 2006. Haughwout et al also showed that these investors defaulted at a much higher rate than single-home owners, representing over 30% of aggregate delinquent mortgage balances. Chinco and Mayer (2011) used a different measure of housing speculators---buyers from out-of-town (i.e., buyers with billing address different from the property address). They showed that in a sample covering all housing transactions of ten metropolitan areas from 2000 to 2008, out-of-town, not local, speculators purchased houses at times and in markets when prices were rising rapidly. They also provided an experimental design to show that out-of-town speculators contributed to excess house price appreciation.

Bubbles in Labs

Smith, Suchanek and Williams (1988) pioneered an experimental approach to study bubbles in laboratories. They created a market for traders to trade dividend-paying assets with a lifetime of a finite number of periods. The only source of asset fundamentals was dividends, whose distributions were publicly announced to all traders. They made the intriguing finding that assets are often traded in high volume at prices substantially above their fundamental values. The same finding was later confirmed by many other studies under a variety of treatments (e.g., Porter and Smith (1995), Lei, Noussair, and Plott (2001), Dufwenberg, Lindqvist, and Moore (2005), Ackert, et al. (2006), Haruvy and Noussair (2006), Haruvy, Lahav, and Noussair (2007), Hirota and Sunder (2007), and Hussam, Porter and Smith (2008)). An important advantage of this experimental approach is that researchers can control the market environment to isolate different mechanisms that drive bubbles. The experimental studies have identified short-sales constraints and investor experience as important factors for the appearance of asset bubbles. Ackert, et al. (2006) and Haruvy and Noussair (2006) find that relaxing short-sales constraints tends to lower prices, while Dufwenberg, Lindqvist, and Moore (2005), Haruvy, Lahav, and Noussair (2007), and Hussam, Porter and Smith (2008), find that as traders gain more trading experience, the divergence in their price expectations is attenuated, and markets become thinner.

II.   Limits of Arbitrage

Before we try to understand how bubbles arise, an immediate question comes to mind: Why don't arbitrageurs eliminate bubbles? In this section I will discuss a set of market and institutional frictions that constrain arbitrageurs from eliminating mispricing in asset bubbles. See Gromb and Vayanos (2010) for an extensive review of this literature.

Short-Sales Constraints

While standard asset pricing theories tend to assume that investors can freely short-sell assets when needed, in practice short sales are often costly, and, in some markets, even impossible. A good example is the housing market, whereby short-selling houses is impractical because different people claim ownership of the same house at the same time often results in legal disputes.[1]

Bris, Goetzmann, and Zhu (2007) provided a thorough account of short sale regulation and practice in the securities markets across 46 countries from January 1990 to December 2001. During this sample period, short sales were prohibited in 10 out of the 46 countries, which included China, Greece, and other less developed countries. In 13 of the 35 countries in which short sales are currently allowed, restrictions existed in 1990 but were lifted at some point during the sample period. This list includes many well developed countries such as Argentina, Finland, Hong Kong, Malaysia, New Zealand, Norway, Spain, Sweden, and Thailand. Interestingly, Malaysia, Hong Kong, and Thailand initially removed restrictions on short selling but later reenacted them. The authors also showed that although short selling is currently legal in most countries, it is only practiced in 25. In some countries, tax rules make shorting very difficult, and in some others (such as Hong Kong) only the largest and most liquid stocks may be shorted.

In most developed countries (such as the U.S., the U.K., and Japan) short-selling stocks and other financial securities is permitted. However, short sales require short-sellers to borrow the securities at a fee in decentralized lending markets before short-selling them. Duffie, Garleanu, and Pedersen (2002) developed a theoretical model to analyze the joint determination of lending fees and asset prices in equilibrium. In their model, agents have different beliefs about an asset's fundamental, and the pessimists need to search for asset owners (who are the optimists in the equilibrium) and bargain over the lending fee. The lending fee is likely to be high if lendable shares of the asset are small and thus difficult to allocate. The lending fee adds to the asset owners' valuation of the asset and thus causes the market price to be even higher than the optimists' beliefs about the asset's fundamental. D'Avolio (2002) and Geczy, Musto, and Reed (2002) empirically analyze the market for borrowing stocks. They find that while it is easy to borrow most stocks at modest fees, the fees can become economically significant when the interest in short-selling becomes high relative to shares available due to either divergence of opinion among investors or special events such as merger announcements that make short-selling profitable. In the

---

[1] Shiller (1998) has been advocating for a long time to create a derivatives market on housing to facilitate short-selling of houses and hedging of housing price risk. He succeeded in persuading Chicago Mercantile Exchange to adopt a set of futures contracts written on Case-Shiller housing price indices of major U.S. cities, but trading volume of these futures contracts remained low even to date due to the lack of liquidity.

sample of stocks analyzed by D'Avolio, the average lending fee for the 9% stocks with the highest fees was 4.3% per annum.

Non-Fundamental Risk

Even in the absence of explicit short-sales constraints, arbitrageurs nevertheless face various forms of non-fundamental risk in arbitraging bubbles (or mispricing in general). DeLong et al. (1990) highlighted the so-called noise trader risk. They argued that when arbitrageurs have a short horizon in an infinite horizon economy, mispricing caused by random demand shocks of noise traders can be self-fulfilling in the equilibrium asset prices even in the absence of any fundamental risk. One can interpret the short horizon of arbitrageurs as a reduced form for explicit financial constraints. The idea is that if arbitrageurs have a one-period horizon and if they expect the price of a long-lived asset with risk-free cash flows to fluctuate with the random demand shocks in the next period, the demand shocks affect the asset price in the current period. This is because arbitrageurs are not able to take long-term positions to trade against the mispricing caused by the demand shocks.

Abreu and Brunnermeier (2002, 2003) labeled the so-called synchronization risk in arbitrage trading. To the extent that arbitrageurs might be informed by the presence of a bubble or mispricing in an asset market at different points of time, the non-common knowledge of their awareness creates a coordination problem in synchronizing their arbitrageur trading. This is because premature attack by an insufficient subset of the arbitrageurs on the bubble is unlikely to succeed and instead exposes them to the risk that the bubble may continue to grow. Anticipating this possibility, each arbitrageur will rationally delay his attack, which in turn allows the bubble to persist for a long time.

Capital Constraints

Arbitrage trading takes capital. Ample empirical evidence suggests that arbitrage capital tends to be specialized and immobile. That is, when one market is short of arbitrage capital, capital does not immediately flow in from other markets. For example, the price of catastrophe insurance tended to increase after major disasters deplete the capital of catastrophe insurers (e.g., Froot and O'Connell (1999)), the convertible bond market became depressed after convertible hedge funds faced large redemption of capital from investors in 2005 (e.g., Mitchell, Pedersen, and Pulvino (2008)), and stock prices tended to display temporary discounts after fire sales by mutual funds (e.g., Coval and Stafford (2007)). See Duffie (2010) for a more extensive discussion of the evidence.

Several frictions can explain such capital immobility. To the extent that arbitrage trading requires market-specific expertise, information barriers (e.g., Bolton, Santos, and Scheinkman (2011)) and search

frictions (e.g., Duffie, Garleanu, and Pedersen (2005)) can both obstruct outside arbitrageurs from freely moving into a capital-constrained market. Even in the absence of these frictions, Shleifer and Vishny (1997) presented another important source of friction that originates from agency considerations. In practice, arbitrageurs tend to trade on others' money by working for hedge funds and financial firms. They thus face agency concerns of their principals (i.e., clients or firm supervisors). That is, after they suffer temporary losses on their arbitrage positions, even if they are certain of the positions' long-run profitability, their principals may doubt their ability and choose to redeem their money. Such redemption forces arbitrageurs to prematurely liquidate positions at losses and thus exacerbates the mispricing. He and Xiong (2010) explicitly analyzed the optimal contracting problem between a principal and a fund manager who faces investment opportunities in multiple markets. They showed, through two insights, that narrow mandates that restrict a manager from investing in a designated market can be optimal despite the apparent investment inefficiency. First, giving a manager more investment flexibility weakens the link between the fund performance and his effort in the designated market, and thus increases agency cost. And, second, the presence of outside assets with negatively skewed returns can further increase the agency cost if the manager has incentive to pursue outside opportunities. These effects motivate narrow mandates and tight tracking error constraints for most fund managers with the exception of those with exceptional talents, and thus provide another explanation for capital immobility.

In the presence of capital immobility, the availability of arbitrage capital in a particular market sector becomes an important determinant of asset prices in the sector. There is a large body of theoretical literature exploring the effects of arbitrage capital on asset price dynamics based on two slightly different approaches (e.g., Xiong (2001), Kyle and Xiong (2001), Gromb and Vayanos (2002), Brunnermeier and Pedersen (2009), Kondor (2009), He and Krishnamurthy (2009), and Danielsson, Shin, and Zigrand (2010)). These models commonly show that after arbitrageurs suffer large losses on their current positions, reduced risk appetite can cause them to liquidate positions despite the fact that the positions become even more profitable. Their liquidation can amplify the effects of the initial shocks. Krishnamurthy (2009) provides a detailed review of such amplification mechanisms.

III.   Heterogeneous Beliefs

Heterogeneous beliefs are a fact of life. It is common for people to take different views on virtually everything, from outcomes of political elections and sports competitions to future inflation and economic growth. In financial markets, it is pervasive for traders to form opposite views about the future performance of stocks and bet against each other. Harris and Raviv (1993) developed a theory of trading

volume in financial markets based on heterogeneous beliefs. The model generated a rich set of empirical predictions on dynamics of trading volume. Kandel and Pearson (1995) provided empirical evidence that supports this theory based on the observed volume-return relation around public announcements.

Figure 6: Disagreement over Inflation Expectations Through Time
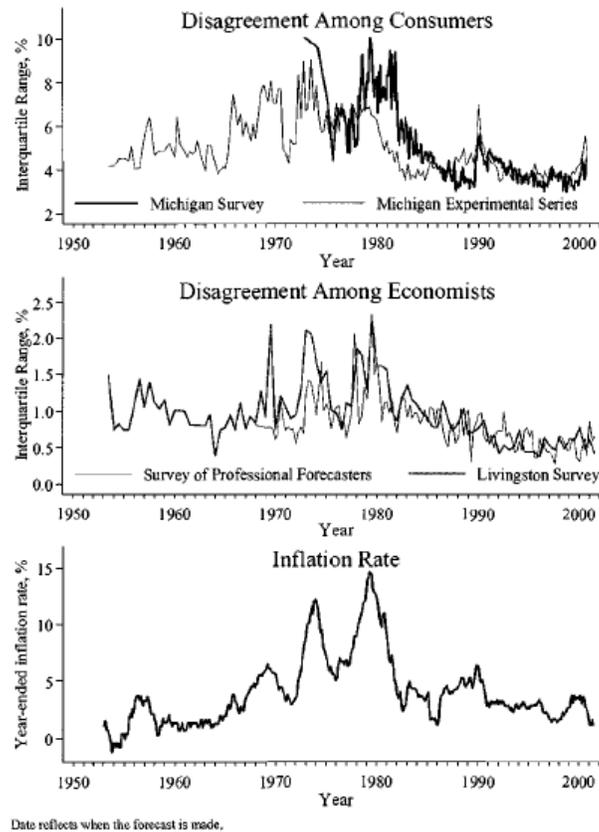
This figure is taken from Figure 3 of Mankiw, Reis, and Wolfers (2004).

It is difficult to directly measure traders' beliefs. Survey data gives a reasonable substitute. Figure 6 (taken from Mankiw, Reis, and Wolfers (2004)) depicts disagreement over inflation expectations, measured by the interquartile range of the individuals' inflation expectations, among economists in the Livingston Survey, among professional forecasters in the Survey of Professional Forecasters, and among individual consumers in the Michigan Survey in 1950-2000. The interquartile range is substantial and varies between 3-8% in the Michigan Survey and between 0.2-2.5% in the surveys of economists and professional forecasters. Interestingly, the disagreement among professional forecasters rises and falls with disagreement among economists and the general public. These series of disagreements clearly indicate significant belief dispersion within groups of people with different backgrounds and are

15

consistent across groups. Welch (2000) surveyed 226 academic financial economists regarding their forecasts of equity premiums and discovered a substantial dispersion in the forecasts. It is now common for academic researchers to use dispersion in surveyed forecasts to measure people's heterogeneous beliefs about various economic variables. For example, Diether, Malloy, and Scherbina (2002), Anderson, Ghysels, and Juergens (2005), and Yu (2011) used dispersion of analysts' earnings forecasts to measure investors' disagreement over individual stocks' fundamentals.

What causes heterogeneous beliefs? Before I discuss the sources, I need to introduce the widely used Bayesian framework, which is the cornerstone for analyzing agents' learning and investment decisions under uncertainty. The standard economics and finance models work as follows. Suppose there is an unobservable fundamental variable θ that determines the payoff of an asset or other relevant state that an agent cares about. Suppose that the agent observes a signal $\tilde{s}$ which is correlated with the fundamental variable θ, and that the agent has some prior beliefs over the joint distribution of θ and $\tilde{s}$. Contingent on observing a particular realization of the signal, the agent uses the Bayes' rule to update his posterior beliefs about θ.[2] If there are multiple agents, one needs to specify each agent's prior beliefs over the joint distribution of θ and signals observed by all agents. Each agent updates his posterior beliefs based on his own signal. If agents end up with different posterior beliefs about θ, they may have started with different prior beliefs about θ, may have observed different signals (i.e., they had asymmetric information), or may have used different updating rules (either because they had different prior beliefs about the joint distribution of θ and the signals or because some of them were irrational). I discuss these sources below.

No Trade Theorem

In the 1970s, the rapidly growing microeconomic literature attempted to link heterogeneous beliefs to agents' asymmetric information. Aumann (1976) clarified an important conceptual point by showing that if agents are rational with a common prior about the joint distribution of θ and the signals, and if there is common knowledge of their posterior beliefs, then their posterior beliefs must be identical. This is because each agent would have used others' beliefs to infer their information and, as a result, their beliefs would have converged. Milgrom and Stokey (1982) and Sebenius and Geanakoplos (1983) extended Aumann's insight to establish the so called no-trade theorem. That is, in the absence of ex ante gains from trade, asymmetric information cannot generate trade among rational agents with a common

---

[2] There is extensive literature analyzing the effects of investor learning on asset price dynamics. See Pastor and Veronesi (2009) for a recent survey.

prior. Taken together, this body of work suggests that asymmetric information has less explanatory power than might have been previously thought.[3]

Heterogeneous Priors

Morris (1995) advocated a view that even rational agents may have heterogeneous prior beliefs. He reviewed different arguments to support the view that the common prior assumption was inconsistent with economists' usual notion of rationality. One of his key arguments built on Savage's (1954) notion of subjective probability: *"Probability measures the confidence that a particular individual has in the truth of a particular proposition, for example, the proposition that it will rain tomorrow. These views postulate that the individual concerned is in some ways 'reasonable', but they do not deny the possibility that two reasonable individuals faced with the same evidence may have different degrees of confidence in the truth of the same proposition."* In fact, the decision theory literature pioneered by Savage treats probabilities separately for individual agents, just like utilities. In particular, it is reasonable for individuals to have heterogeneous prior beliefs about new things, such as technological innovations and IPOs, because they don't have any useful information to form any prior belief. In a sense, economics does not provide much guidance on how individuals form their prior beliefs. Prior beliefs probably depend on individuals' background and experience. Economists tend to treat people's prior beliefs as part of their preferences.

If heterogeneous beliefs are derived from heterogeneous priors, one may argue that as individuals obtain sufficient information over time, learning will eventually cause their beliefs to converge. While appealing, this argument does not always hold true. Endogenous learning explains one reason. In the multi-armed bandit problem studied by Rothschild (1974), a gambler chooses repeatedly between two slot machines in a casino, one with a known probability of payout and the other with an unknown probability. The optimal strategy might allow experimentation with the unknown machine. But it might also require abandoning the unknown machine forever after some finite number of trials if it does not perform well. The gambler thus may never learn the true probability of success on that machine. This shows that heterogeneous beliefs can persist in an environment where learning is costly and endogenous.

Even in the absence of endogenous learning, the eventual convergence of agents' heterogeneous beliefs is not guaranteed. Kurz (1994) defined a belief to be rational if it generates the same long-run empirical frequencies as the data. In a stationary economic system there is a unique rational belief. In

---

[3] While asymmetric information alone cannot lead to trade, a large literature analyzes liquidity and trading in settings with both asymmetric information and random supply shocks following the classic analysis of Grossman and Stiglitz (1980) and Kyle (1985). The presence of random supply shocks creates ex ante gains from trade for rational speculators. The recent work of Banerjee (2011) contrasts implications of asymmetric information and heterogeneous prior beliefs for trading volume and asset price dynamics.

contrast, Kurz pointed out that if the system is not stationary there typically will be many rational beliefs. Even in a stationary economic system, learning requires that agents know the conditional distribution of signals given the fundamental variable. Acemoglu, Chernozhukov, and Yildiz (2009) showed that when agents are uncertain about the signal distributions, even vanishingly small individual uncertainty about the signal distributions can lead to substantial (non-vanishing) differences in asymptotic beliefs.

Overconfidence

Extensive evidence shows that excessive trading severely undercuts portfolio performance of individual investors in different countries, e.g., Odean (1999), Barber and Odean (2000), Grinblatt and Keloharju (2000), and Barber, Lee, Liu, and Odean (2009). For example, Barber and Odean (2000) analyzed trading records of a large sample of retail investors in the U.S. and found that trading caused the returns of the most active 20 percent of investors to under-perform the market returns by over 5 percent per year. While many reasons could induce investors to trade, such as heterogeneous priors, it is difficult to fully explain such poor trading performance without relating it to certain psychological biases.

A large body of behavioral finance literature highlights the finding that people suffer from a range of well-established psychological biases, such as overconfidence, limited attention, representativeness and conservatism, in making financial decisions. See Hirshleifer (2001) and Barberis and Thaler (2003) for extensive reviews of the literature. In particular, extensive psychology literature on calibration showed that people tend to believe their knowledge is more accurate than it really is. In a classic experiment, Alpert and Raiffa (1982) asked individuals to give confidence intervals for predicting extreme events. Their 98% confidence intervals were very narrow and contained the true quantity only 60% of the time. Experts can be more prone to overconfidence than non-experts when predictability is low and evidence is ambiguous (Griffin and Tversky (1992)). Consistent with this, a recent survey of CFO predictions of S&P stock returns by Ben David, Graham, and Harvey (2010) showed that the realized returns were within the CFOs' [10%, 90%] intervals only 33% of the time.

Overconfidence causes agents to exaggerate the precision of noisy signals and thus to over-react to the signals. When agents over-react to different signals, they may end up with substantially different beliefs. Kyle and Wang (1997), Odean (1998), and Scheinkman and Xiong (2003) derived models with heterogeneous beliefs originated from agents' overconfidence. By using very different model frameworks, they highlighted that such heterogeneous beliefs can lead to excessive trading and asset price bubbles. In particular, Scheinkman and Xiong presented a dynamic framework in which two overconfident investors are symmetrically overconfident about two different public signals. Their overreactions to the distinct signals cause their beliefs to diverge, but their unbiased learning through other signals gives another force

for their beliefs to converge over time. Interestingly, in each investor's mind, the other agent's belief converges to his according to a linear mean-reverting process. This stationary belief structure makes it convenient to analyze dynamic asset market equilibrium with heterogeneous beliefs, which I will discuss in the following sections.

Distorted Information Transmission

Another important source of heterogeneous beliefs is biases in information transmission. In practice, investors heavily rely on financial advisors and analysts for information regarding investment decisions. In working for their principals (investors), financial advisors face incentive problems and reputation concerns, which may end up distorting their investment recommendations (e.g., Lin and McNichols (1998) and Hong and Kubik (2003)). Smart investors recognize such distortion and can discount the recommendations to de-bias the transmitted information. However, other investors may be too naïve to recognize such distortions (e.g., Malmendier and Shantikumar (2007)). As a result, heterogeneous beliefs may arise through biases in information transmission between investors and financial advisors. Hong, Scheinkman and Xiong (2008) highlighted such a mechanism. At times of technological innovations, financial advisors face reputation concerns among investors that they may be old fogies and may be downwardly biased against new technologies. In order to signal their types, even a well-intentioned financial advisor may choose to upwardly bias his recommendation as a way to signal his type as a tech-savvy. His biased recommendation thus stimulates disagreements among smart investors who de-bias the recommendation and naïve investors who do not.

IV. Resale Option Theory of Bubbles

Harrison and Kreps (1978) suggested that in a dynamic environment with time-varying heterogeneous beliefs and short-sales constraints, an asset buyer may be willing to pay more than his own expectation of the asset's fundamental. This is because he holds the option of reselling the asset to other more optimistic buyers for a speculative profit in the future. This notion of resale options provides a powerful mechanism for analyzing speculative behavior and asset bubbles. In this section, I review a number of studies that exploit this insight. I discuss the implications of heterogeneous beliefs in general equilibrium settings without short-sales constraints in Section VI.

Static Disagreement Models

I first start with the joint effects of heterogeneous beliefs and short-sales constraints in static settings. Miller (1977) argued that short-sales constraints can cause stocks to be overpriced when investors have

heterogeneous beliefs about stock fundamentals. In the presence of short-sale constraints, stock prices reflect the views of the more optimistic participants. Since the pessimistic investors are not allowed to short sell, prices in general will be higher than what would prevail in the absence of short-sale constraints. That insight motivated a series of studies to analyze effects of heterogeneous beliefs on stock over-valuation. Jarrow (1980) extended the setting to have multiple assets and mean-variance preference. He showed that whether over-valuation occurs depends on individuals' beliefs about the covariance matrix of assets' payoffs.[4]

Despite of the subtlety in the theoretical arguments, there is strong evidence supporting overvaluation caused by heterogeneous beliefs and short-sales constraints. By extending Miller's model, Chen, Hong and Stein (2002) established breadth of a stock's ownership as a novel measure of whether short-sales constraints were binding for the stock. That is, when breadth is low (i.e., when few investors have long positions), the short-sales constraints are likely to be binding and the stock's price tends to be high relative to its fundamental. Consistent with this model prediction, they found that stocks whose change in breadth in the prior quarter was in the lowest decile of the sample significantly underperformed those in the top decile. Diether, Malloy, and Scherbina (2002) directly used dispersion in analysts' earnings forecasts to measure investors' heterogeneous beliefs about stocks' fundamentals. They found that stocks with higher forecast dispersion earned significantly lower future returns than otherwise similar stocks. Yu (2011) aggregated individual-stock analyst forecast dispersions to measure portfolio disagreement and found that a higher market disagreement predicts poor future market return.

Dynamic Disagreement Models

In dynamic settings with time-varying heterogeneous beliefs, an asset owner has a speculative motive to resell his asset holding to an even more optimistic buyer at a speculative profit (Harrison and Kreps (1978)). This motive leads the asset owner to value the asset at a price higher than his already optimistic belief and thus forms a legitimate bubble under any restrictive definition of bubbles. This motive also leads to over-trading and rich implications about several aspects of bubbles.

I illustrate the basic idea through an example with two periods and three dates: $t = 0, 1, 2$. Two risk-neutral investors, A and B, trade a risky asset. The asset gives a final payoff on date 2, which can take three possible values 100, 50, or 0. The realization of the final payoff depends on the evolution of a binomial tree, depicted in Figure 7. The tree can either move up or down in each of the two periods from

---

[4] Gallmeyer and Hollifield (2008) analyzed the effects of a market-wide short-sale constraint in a dynamic economy with heterogeneous beliefs. They showed that imposing the short-sale constraint may or may not increase the stock price depending on investors' intertemporal elasticity of substitution due to the accompanying effect on the equilibrium interest rate.

date 0 to date 1 and from date 1 to date 2. There are thus two intermediate states on date 1, which I denote by u after an up move from date 0 and d after a down move. On date 2, the final payoff is $100 after two consecutive up moves, $50 after a combination of one up move and one down move, and $0 after two consecutive down moves. The move of the tree is random and has independent and identical distributions over the two periods. The probability of an up movement π determines the asset's fundamental and is unobservable. I denote investor i's belief in state j ($j \in \{0, u, d\}$ where 0 refers to date 0) by π_j^i. Suppose that the two investors have the following state-dependent beliefs:

$$\pi_0^A = \pi_u^A = \pi_d^A = 0.5, \pi_0^B = 0.5, \pi_u^B = 0.8, \pi_d^B = 0.2.$$

In this specification, investor A holds a constant belief of 0.5 across date 0 and the two states on date 1, while investor B's belief fluctuates from 0.5 on date 0 to 0.8 in state u of date 1 and 0.2 in state d. One can interpret the fluctuation in investor B's beliefs as a result of his learning. For simplicity, I skip modeling the learning process and directly state the beliefs. I also assume that the interest rate is zero and short-sales of the asset are not allowed.

It is straightforward to compute the two investors' fundamental valuations of the asset. In state u of date 1, investor A's fundamental valuation is $75 while investor B's is $90; in state d of date 1, investor A's fundamental valuation is $25 while investor B's is $10; on date 0, both investors A and B have the same fundamental valuation of $50.

Figure 7: An Illustration of the Resale Option Theory of Bubbles
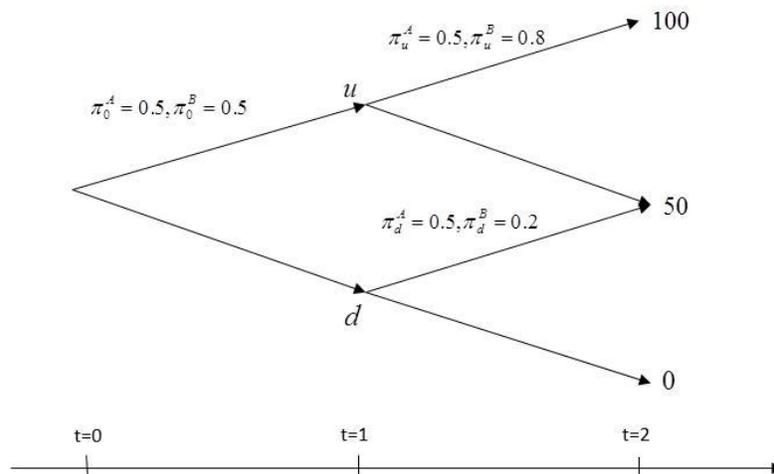
How much is the market valuation of the asset? In state u of date 1, investor B has a higher valuation of $90 and would thus choose to buy the asset. Investor A would sell but cannot sell more than

what he owns due to short-sales constraints. Suppose that investor B has sufficient cash and can bid up the price to his valuation. I will discuss issues that might occur when the optimist is cash-constrained in the next section. I also assume that the market is competitive (i.e. there are multiple investors like B) and would push the price to the buyer's reservation valuation. Thus, investor B buys the asset from investor A at a market price of $90. Note that the short-sales constraints bias the price to the optimist's valuation, as argued by Miller (1977). If the pessimist were allowed to short sell the asset, his short selling can neutralize the optimist's optimism. Similarly, in state d of date 1, investor A buys the asset from investor B at a market price of $25. By backward induction, on date 0 both investors value the asset at $57.5 even though their expectations of the asset's final payoff are both $50! The difference of $7.5 reflects the value of the asset owner's resale option. That is, there is a possibility for the owner to sell the asset to the other investor at a price higher than the owner's own valuation of the asset on date 1. From the perspective of investor A, he expects to sell the asset to investor B in state u of date 1 at a price of $90, which is $15 higher than his own fundamental valuation at the time, and continue to hold the asset if state d occurs. The option to resell in state u gives investor A the extra $7.5 valuation on date 0. Similarly, investor B expects to sell the asset to investor A in state d of date 1 at a price of $25, which is $15 higher than his fundamental valuation, and continue to hold the asset if state u occurs. The option to resell in state d also gives investor B the extra $7.5 valuation on date 0. This resale option reflects the speculative behavior highlighted by Harrison and Kreps (1978). Its value contributes to a non-fundamental component in asset prices that is in excess of the fundamental valuation of the asset owner who already holds the more optimistic belief in the market.

Short-sales constraints are important although not essential for an asset owner to realize his resale profit. One can relax the restrictive short-sales constraints in the example and allow more modest but costly short sales. The asymmetry between optimists and pessimists' willingness to take long versus short positions would still allow an asset owner to profit from reselling the asset to future optimists. As a result, his resale option is still valuable and would thus motivate him to value the asset at a price higher than his expectation of the asset's fundamental.

In this example, I take the two investors' state-dependent beliefs as given. Morris (1996) provided a more general model in which agents start with heterogeneous prior beliefs regarding the fundamental of an IPO and update their beliefs based on a stream of public information over time. Their beliefs fluctuate over time in response to the information flow. In particular, their beliefs may cross each other even though the beliefs eventually converge in the long run. The fluctuation induces the agents to engage in speculative trading against each other and to value the IPO at prices higher than any agent's fundamental valuation. Morris used this model to explain the widely observed overvaluation of IPOs.

Characterizing Price Bubbles

Scheinkman and Xiong (2003) demonstrated that a modest amount of heterogeneous beliefs is sufficient to generate a substantial price bubble through a recursive structure of resale options and that this mechanism can lead to the joint occurrence of asset price bubbles and trading frenzies, a common characteristic of many historical bubble episodes. In their model, heterogeneous beliefs arise from agents' overconfidence and follow a stationary, linear structure. There are two agents who trade a risky asset with an unobservable fundamental. The asset's fundamental follows a linear diffusion process with constant volatility. The agents infer the value of the asset's fundamental based on observing the asset's dividend flow and two streams of signals. Each agent has a distinct, favorite signal and is overconfident about its precision, although he is objective about the information contained in the other signal and the dividend flow. Overconfidence causes the agents to overreact to their favorite signals. Their overreactions cause their beliefs to diverge from each other, while their consistent reactions to the dividend information cause their beliefs to converge over time. From each agent's perspective, the other agent's belief mean reverts to his own according to a linear diffusion process. This linear belief structure together with a linear structure of the asset's fundamental makes it possible to decompose an asset owner's valuation of the asset as the sum of his expectation of the asset's fundamental and a resale option whose value is determined by the difference of the two agents' beliefs. The authors formulated this resale option component to study asset price bubbles.

The asset owner faces an optimal stopping problem in deciding whether to exercise his resale option and resell the asset to the other agent. In equilibrium, he sells the asset once the difference in beliefs from his perspective passes an optimal threshold. That is, a trade occurs when the buyer's belief surpasses the owner's by a margin determined by trading cost and the owner's option value of holding onto to the asset. Interestingly, this resale option component has a recursive structure in the sense that when the current asset owner resells the asset, the buyer obtains another resale option. This recursive structure implies that the bubble component can be substantial even when small differences in beliefs are sufficient to generate a trade. The authors show that as the difference of agents' beliefs becomes more volatile (due to either more pronounced overconfidence or a greater volatility of the asset's fundamental) the agents tend to trade more frequently and the resale option becomes more valuable. As a result, a greater asset bubble is accompanied by a more intensive trading frenzy.

This analysis offers several useful results: First, it is not necessary to require an outrageous amount of optimism from all investors in order to generate a large bubble; even in the absence of any aggregate belief distortion of all investors, fluctuations of heterogeneous beliefs among them can nevertheless lead

to significant price bubbles. This condition is rather modest and realistic based on my earlier discussion regarding the sources of heterogeneous beliefs. Second, the bubble component derived from the resale option theory has stationary time-series properties, in sharp contrast to the explosive rational bubble of Blanchard and Watson (1983), which is based on a framework with homogeneous agents. Third, the intensity of trading reflects the fluctuations of agents' heterogeneous beliefs and is thus correlated with the resale option component in the asset price.

These results are useful for understanding two common characteristics of the historical bubbles discussed in Section I: 1) bubbles tend to coincide with technological and financial innovations; and 2) bubbles tend to coincide with frenzied trading. As new technology and new financial innovation usually arrive with great fundamental uncertainty, they not only stimulate heterogeneous prior beliefs among investors but also create an unexplored environment for overconfidence which affects investors and causes them to overreact to their favorite signals. Thus, it is natural to observe bubbles coinciding with innovations. The logic for the joint occurrence of bubbles and frenzied trading is evident from my earlier discussion.

There is even systematic evidence of the joint occurrence of high prices and overtrading in both cross-sectional and time-series analyses. Hong and Stein (2007) provided evidence that in a sample of 1000 largest stocks in the CRSP database from 1986 to 2005, glamour stocks (i.e., stocks with high market-to-book ratios) had higher turnover rates than low-priced value stocks and that this difference was particularly pronounced during the Internet bubble. They also represented a significant correlation of 0.49 between annual S&P return and annual change in NYSE turnover from 1901 to 2005. In a different sample, Mei, Scheinkman, and Xiong (2009) analyzed the price ratio of domestic and foreign shares issued by the same firms in China's stock market. As these shares are entitled to the same cash flow and voting rights, their price ratios gave a measure of the bubble component caused by the domestic investors' speculative behavior. Consistent with the model's prediction, they found a significant, positive correlation between the price ratio and the domestic shares' turnover rate across different pairs of domestic and foreign shares. Furthermore, Xiong and Yan (2011) found that during the Chinese warrants bubble, the higher priced warrants also tended to have higher turnover rates.

Given the severe economic consequences of bubbles, policy makers and investment practitioners face the challenge of how to detect an ongoing bubble before it collapses. The ability to do so not only helps investors and financial institutions avoid losses from investing in bubbles but, more important, allows policy makers to more efficiently control risk taken by financial institutions and prevent formation of systemic risk. The joint occurrence of high prices and frenzied trading offers a useful hint at how to

design such a detection mechanism. Based on the evidence discussed above, it works well in sample, although its out-of-sample predictive power remains to be examined.

Scheinkman and Xiong (2003) also examined the effect of Tobin's transaction tax on asset price bubbles. Interestingly, a transaction tax can substantially reduce the amount of trading in markets with small transaction costs but would have a limited effect on the size of the bubble or on price volatility. This is because in response to the increased trading cost each agent would mitigate its impact by raising the selling threshold in difference of beliefs to a level higher than necessary to offset the trading cost. The increased selling threshold can dramatically reduce the trading frequency but at the same time internalize the impact of the increased trading cost on the asset price. This analysis thus casts doubt on the effectiveness of Tobin's transaction tax on curbing asset bubbles. It also suggests that asset bubbles can arise even in a market with high transaction costs such as housing markets.

Hong, Scheinkman and Xiong (2006) highlighted the role of asset float in bubble formation and burst. The authors presented a discrete-time model with finite periods, in which investors with limited risk-bearing capacities trade an asset with a finite float (i.e., number of shares outstanding). The investors' time-varying heterogeneous beliefs together with short-sales constraints motivate them to value the resale option associated with owning the asset. Their limited risk-bearing capacities render the asset float an important determinant of the value of the resale option. A larger float mean that it takes a greater divergence of opinion in the future for the current asset owner to resell the asset at a speculative profit, and thus makes the resale option less valuable. This simple insight can help explain the dramatic deflation of Internet stocks in spring 2000 when the tradable shares of a large number of Internet firms largely increased after their lockup restrictions expired and insiders became free to sell their stock holdings, as shown by Ofek and Richardson (2003). It also rationalizes the incentive of the South Sea Company to lobby for the passage of the Bubble Act during the South Sea Bubble. Their model also shows that the bubble component of the asset's price increases as the asset's maturity approaches. This is because there is less time left for an asset owner to resell his holding. This prediction is supported by the gradual, decreasing trend observed in the Chinese warrants bubble (Xiong and Yu (2011)), which sharply contrasts to the explosive bubble paths that are commonly perceived.

Crashes of Bubbles

What causes bubbles to crash? During bubbles it is difficult to precisely determine assets' fundamentals. It is reasonable to expect that as more information becomes available over time, the uncertainty would gradually diminish and market participants would eventually recognize overvaluation of asset prices relative to their fundamentals. As a result, asset price bubbles would eventually deflate.

However, as I discussed earlier, historical asset bubbles often burst at unpredictable times and over brief periods. It is thus difficult to explain the intensity of bubble burst simply using market participants' gradual learning process.

The existing economics literature offers several additional factors that might trigger crashes of bubbles. Hong and Stein (2003) provided an interesting insight based on the revelation of previously hidden negative information after a market downturn in an environment with heterogeneous beliefs and short-sales constraints. Due to short-sales constraints, pessimistic investors do not initially participate in the market and their information is not revealed in prices. However, if other previously optimistic investors bail out of the market, the originally pessimistic investors may become the marginal buyer to support the market. Their willingness to step in reveals their previously hidden information. A rational arbitrageur may infer the hidden information to be more negative than previously thought if the originally pessimistic investors fail to give sufficient support for the market. Such an inference amplifies the initial price drop and can thus lead to a market crash. This theory helps explain sudden market crashes in the absence of dramatic news events.

Abreu and Brunnermeier (2003) used a generalized notion of heterogeneous beliefs to model the synchronization problem of rational arbitrageurs in coordinating their selling strategies during a bubble. In their model, a bubble caused by irrational investors emerges at a random time and then grows at an exogenous speed. The model focuses on a group of rational arbitrageurs who are allowed to short-sell the bubble. The bubble will eventually burst under sufficient selling pressure from arbitrageurs. Arbitrageurs are privately informed of the presence of the bubble at different times. That is, when an arbitrageur gets informed of the bubble, he faces the uncertainty that other arbitrageurs may not yet be informed. The heterogeneity in the times that arbitrageurs get informed leads to heterogeneous beliefs about the presence of the bubble at a given time, which, in turn, makes it difficult for arbitrageurs to coordinate their short-selling of the bubble. As riding the bubble before it crashes is profitable, it is rational for each arbitrageur to initially ride on the bubble rather than short-sell. Thus, the bubble may persist for a substantial time even though arbitrageurs are well informed. The model also shows that news events, by enabling synchronization of arbitrageurs' selling, can have a disproportionate impact relative to their intrinsic informational content in driving the crash of the bubble.

Bubbles and Overinvestment

A severe consequence of asset bubbles is firms' value-destroying activity such as overinvestment. The investment boom in the telecom industry during the Internet bubble is a good example. There is also growing evidence of inefficient stock markets affecting firms' investment decisions (e.g., Morck, Shleifer

and Vishny (1990), Blanchard, Rhee and Summers (1993), Stein (1996), Baker, Stein and Wurgler (2003), Panageas (2004), Gilchrist, Himmelberg and Huberman (2005), and Polk and Sapienza (2009)). In particular, Gilchrist, Himmelberg and Huberman (2005) provided evidence that firm investment increases with investors' heterogeneous beliefs measured by dispersion in stock analysts' earnings forecasts. These papers tend to emphasize that when stocks are overvalued, firms overinvest by taking advantage of a cheap source of capital. However, this argument leaves open an important issue of why firms run by managers on behalf of their investors would engage in inefficient investment behavior that is detrimental to their investors' interests. This issue lies at the core of understanding shareholder value and corporate governance in an inefficient market environment. In the aftermath of the collapse of the Internet bubble, many pundits pointed out how executives and directors of many companies managed to enrich themselves by selling their shares shortly before their company's stock price crumpled, and raised severe concerns about corporate governance of these companies (e.g., Bebchuk and Fried (2004)).

Bolton, Scheinkman and Xiong (2005, 2006) addressed this issue by providing a theory of optimal executive compensation and short-termist behavior in a speculative environment in which investors hold heterogeneous beliefs about a firm's fundamental value and face short-sales constraints. Following the aforementioned resale option theory of bubbles, in this market environment the firm's current share price contains not only the firm's long-run fundamental value but also the resale option component. That is, the current shareholders may be able to sell their shares to other more optimistic investors in the future at a price higher than their own fundamental valuation. This in turn implies that it may be in the interest of the current shareholders to pursue short-termist strategies that increase the firm's resale option value at the expense of its long-run fundamental. Thus, even in the absence of any governance failure, the current shareholders may choose an optimal compensation contract with an emphasis on short-term stock performance in order to motivate the firm executive to pursue short-term strategies, such as overinvesting in Internet technology, which maximize the firm's market share price rather than its long-run fundamental. In contrast to the rent seeking view of executive compensation, the theory of Bolton, Scheinkman and Xiong interpreted the short-term stock-price-based executive compensation as a tool for the current shareholders to exploit the mispricing of future investors.

Other Bubble Theories

Over the years economists have developed different theories to explain asset price bubbles, which include the rational bubble theory, the agency based bubble theory, and the behavioral based feedback loop theory. I briefly discuss these theories here.

Blanchard and Watson (1983) provide a rational bubble model that is fully consistent with rational expectations and constant expected returns. They use a discrete-time setting with homogenous rational investors and infinite periods, and specify the price of an asset with two components--a fundamental value component and a rational bubble component. The fundamental component is determined by the asset's discounted cash flow. The rational bubble component is independent of the asset's fundamental and fluctuates over time on its own. As long as it grows on average at the same rate as the discount rate, it is consistent with rational expectations. Blanchard and Watson allow the bubble component to burst with a constant probability in any period. If it does not burst, it grows a rate higher than the discount rate. Rational bubbles have attracted considerable attention in the academic literature. However, there are both theoretical and empirical arguments that can be used to rule out the existence of such a rational bubble component in asset prices. In particular, the asset must be infinitely lived and there cannot be any upper limit on its price so that its price can be expected to grow at the discount rate. These restrictions prevent rational bubbles from appearing in prices of assets with finite maturities such as warrants and bonds. See Campbell, Lo and MacKinlay (1997, Chapter 7.1) for an excellent summary of other arguments. In contrast to the explosive time-series properties of the rational bubble, the bubble component generated by the resale option does not require an infinite maturity and is stationary over time.

The behavioral finance literature suggests that various behavioral biases, such as representativeness bias and self-attribution bias, can lead individual investors to positively feed back to past returns (e.g., Barberis, Shleifer and Vishny (1998), Daniel, Hirshleifer and Subrahmanyam (1998), and Gervais and Odean (2001)). Motivated by these studies, Shiller (2005) advocated a feedback loop theory of bubbles. In this theory, "*Initial price increases caused by certain precipitating factors lead to more price increases as the effects of the initial price increases feedback into yet higher prices through increased investor demand. This second round of price increase feeds back again into a third round, and then into a fourth, and so on. Thus the initial impact of the precipitating factors is amplified into much larger price increases than the factors themselves would have suggested.*" There is ample empirical evidence, such as the presence of short-run momentum and long-run reversals of stock prices, which support the feedback loop theory of bubbles. See Hirshleifer (2001) and Barberis and Thaler (2003) for complete reviews of the related empirical literature. It is useful to note the sharp contrast between the feedback loop theory and the resale option theory of bubbles. The former requires substantial aggregate belief distortions of investors in order to generate a significant price bubble, while the latter emphasizes that even in the absence of any large aggregate belief distortion, a sizable bubble could arise through a modest amount of time-varying heterogeneous beliefs.

Allen and Gorton (1993) and Allen and Gale (2000) developed models to show that bubbles can arise from agency problems of institutions. Allen and Gorton (1993) showed that in the presence of asymmetric information and contract frictions between portfolio managers and investors that hire them, managers bear limited downside risk because the worst that can happen to them is that they get fired. As a result, they have incentives to churn and seek risk at the expense of their investors. Allen and Gale (2000) analyze the risk-shifting incentive of investors who use borrowed money from banks to invest in risky assets and who can avoid losses in low payoff states by defaulting on the loan. In both of these models, assets can trade at prices that do not reflect their fundamentals. As pointed out by many pundits (e.g., Rajan (2010)), these incentive problems are highly relevant in understanding the recent financial crisis of 2007-2008. However, distorted beliefs and heterogeneous beliefs probably played equally important roles as distorted incentives in contributing to the boom that preceded the crisis (e.g., Benabou (2009) and Barberis (2011)). It will probably take years for economists to fully separate the roles played by these different factors.

V.   Credit Cycles

Hyman Minsky and Charles Kindleberger have long pointed out the importance of credit market expansion and tightening as key factors driving the boom and bust of financial markets as well as the real economy. Economists have a renewed interest in credit cycles in the aftermath of the recent financial crisis in 2007-2008. Adrian and Shin (2010) provided vivid empirical evidence that marked-to-market leverage of financial intermediaries is procyclical and that changes in the aggregate balance sheets of intermediaries forecast changes in financial market risk and liquidity. There is a growing body of theoretical literature analyzing the effects of collateral constraints and margin constraints on asset price dynamics. Kiyotaki and Moore (1997) developed a dynamic model to show that through their role as collateral for loans, durable assets (such as land) can amplify small, temporary shocks to technology or income distribution to generate large, persistent fluctuations in output and asset prices. Gromb and Vayanos (2002) showed that margin constraints can force arbitrageurs to prematurely liquidate arbitrage positions at losses and thus contribute to inefficiency in asset prices. Brunnermeier and Pedersen (2009) highlighted that margin constraints can be destabilizing and market liquidity and traders' funding liquidity can be mutually reinforcing, leading to liquidity spirals. In other words, deterioration in market liquidity tightens traders' margin constraints and thus funding liquidity, which in turn further reduces market liquidity. Garleanu and Pedersen (2011) developed a dynamic equilibrium model to show that margin constraints can lead to deviations from the law of one price.

A number of commentators including Shiller (2000) and Reinhart and Rogoff (2009) identified the optimism of a set of investors as an important driving force in the recent housing market boom as well as many other historical financial market booms, which all ended with dramatic financial crises due to the leverage used by these investors. Geanakoplos (2003, 2010) developed an elegant framework with multiple periods and heterogeneous beliefs to analyze the joint dynamics of assets' collateral values and market values. This framework features a risky asset and heterogeneous agents with a continuum of beliefs regarding the asset's fundamental. The agents with more optimistic beliefs would like to acquire the asset but are cash constrained. They can use the asset as collateral to raise debt financing from other agents. In the equilibrium, the credit market and asset market boom together: on one hand, a higher collateral value allows an optimistic buyer to obtain financing at a lower cost and thus to bid up the asset price; on the other, a higher asset price increases the asset's collateral value. After a bad fundamental shock in the future, the credit market equilibrium exacerbates the impact of the shock on the asset market equilibrium. In addition to the deterioration in the asset's fundamental value, the shock wipes out the initial asset owners who have used maximum riskless debt to finance their asset positions and thus shifts down the clientele of the marginal asset owner. The shock also reduces the asset's collateral value and thus further amplifies the effect of the shock on asset prices. Fostel and Geanakoplos (2008) extended this framework to an international setting and showed that leverage cycles can cause several interesting phenomena such as contagion, flight to collateral, and issuance rationing and explain the volatile excess of emerging economies to internal financial markets. Cao (2011) developed a general equilibrium model with heterogeneous beliefs and collateral constraints to analyze the effects of collateral constraints on asset price volatility and real investment volatility.

I will use a simple example to illustrate the key mechanism. Suppose that there are two types of agents, A and B, trading a risky asset in a setting with three dates $t = 0, 1, 2$. I use the same binomial tree used in the previous example for illustrating the resale option theory of bubbles in Section IV to represent the asset's fundamental in this example. Like before, short-sales of the asset are not allowed and the interest rate is normalized to zero. I will also use the same notation. The asset's liquidation value on date 2 can be 100, 50, or 0. The type-A and type-B agents hold heterogeneous beliefs about the probability of the tree moving up in each period. Instead of allowing their beliefs to change over time, this example restricts the beliefs of the two types of agents to be constant: $\pi^A = 0.5$ and $\pi^B = 0.8$. Figure 8 depicts the binomial tree and the agents' beliefs. It is easy to compute the agents' fundamental valuations of the asset in different states. In state u of date 1, type-A agents' fundamental valuation is \$75 while type-B agents' is \$90; in state d of date 1, type-A agents' fundamental valuation is \$25 while type-B agents' is \$40; on date 0, type-A agents' fundamental valuation is \$50 while type-B agents' is \$80.

Figure 8: An Illustration of Credit Cycles
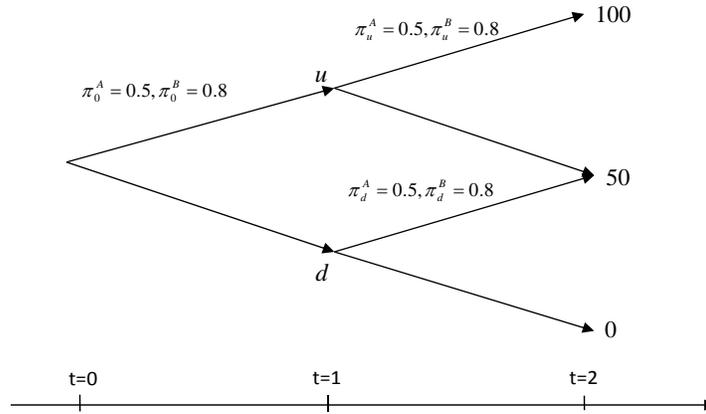
Suppose that on date 0 type-A agents initially own all of the asset, which is normalized to one share and that there is one unit measure of type-B agents, each with $40 in cash endowment. Given the optimism of type-B agents, they desire to acquire the asset, but they are cash constrained. How many shares of the asset can type-B agents buy and at what price? This depends on the amount of financing they can obtain. Suppose that they can use the asset as collateral to raise debt financing from type-A agents on date 0. The collateral serves to mitigate the default risk of the borrower. That is, if the borrower fails to pay back the promised payment, the lender will seize the collateral. Thus, the quality of the collateral determines the default risk of the debt.

Suppose that the asset price on date 0 is between the type-A and type-B agents' fundamental valuations. Then, an individual type-B borrower needs to structure his debt financing to maximize the expected total return on his investment. On one hand, a greater initial credit allows him to establish a large position; on the other, he is concerned by the cost of the credit. In particular, the heterogeneous beliefs between type-A and type-B agents imply that a type-A lender is more concerned by the downside risk of the asset than a type-B borrower. As a result, if the debt is risky, the type-A lender will demand a default premium higher than the default risk assessed by the type-B borrower. Geanakoplos (2010) argued that this wedge motivates the type-B borrower to use the risk-free debt that gives the maximum amount of initial credit, which, in this example, means a one-period debt with a promise to pay $25 in period 1. This debt contract is risk-free as even if the borrower fails to pay back the debt in the lower state d of date 1, the type-A lender can always seize the asset and hold it at his own fundamental valuation of $25. Through this debt contract, type-B agents can obtain $25 of initial credit on date 0 from type-A agents to acquire the asset. Together with type-B agents' initial cash endowment of $40, the

initial credit allows them to bid up the asset price to $65 per share, a value between the type-A and type-B agents' fundamental valuations. In this sense, the asset's collateral value determines its market price. If the lower state d realizes on date 1, type-B agents are all wiped out as their asset holding is just enough to pay back their debt. As a result, the asset price in state d falls to the fundamental valuation of type-A agents, $25, rather than somewhere between the type-A and type-B agents' valuations. In this sense, the credit cycle forces a fire-sale of the asset and, as a result, amplifies the impact of the negative shock on the asset price.

In this example, agents' beliefs stay constant over time. In a more general setting with time-varying beliefs, an individual type-B agent's investment and financing decisions are more subtle. He can use either short-term or long-term debt to finance an asset acquisition; he can also save cash to wait for the fire-sales of other optimists in the intermediate lower state. The last choice is particularly attractive if the heterogeneous beliefs between the pessimists and optimists widen in this state, which makes the price discount in the fore sale particularly steep. He and Xiong (2012) provided a model with many periods and time-varying heterogeneous beliefs to analyze the equilibrium debt financing used by optimists. They showed that an individual optimist either use risk-free one-period debt to finance his asset acquisition or save cash for future fire-sales depending on the dynamics of agents' heterogeneous beliefs and the optimists' initial cash endowment. They also derived a risk-neutral representation of the equilibrium asset price and prices of different debt contracts collateralized by the asset. The model also highlights that the asset's marketability (i.e., whether there exists a market to trade the asset in the intermediate dates) can have a significant impact on the asset's collateral value and thus the asset price.

As I discussed in Section IV, Miller (1977) posited that in the presence of short-sales constraints, greater belief dispersion leads to greater overvaluation because asset price is determined by optimists. Simsek (2010) re-examined this argument in a setting where the optimists are cash constrained and need to obtain collateralized debt financing from the pessimists. He showed that the distribution of agents' disagreements matters: if agents disagree only about the probability of upside states, a greater disagreement leads to a higher asset price, just as posited by Miller; if agents disagree only about the probability of downside states, a greater disagreement leads to a lower asset price because a greater disagreement about downside states increases the optimists' borrowing cost from the pessimists.

VI.   General Equilibrium Models with Heterogeneous Beliefs

Even in the absence of short-sales constraints and leverage constraints, heterogeneous beliefs can lead to rich implications for asset price dynamics in general equilibrium settings. As the standard

representative-agent based models treat agents as homogeneous, they face several well-known challenges in fitting the data. In these models, the aggregate fundamental fluctuations are sufficient to measure the risk faced by individual agents. The observed asset returns appear excessively volatile relative to the volatility of aggregate fundamentals (e.g., Shiller (1981)). These models also require agents to have implausibly high risk aversion to match the observed equity premium (e.g., Mehra and Prescott (1985)). Heterogeneous beliefs cause agents to speculate against each other, which, in turn, lead to endogenous fluctuations in individual agents' wealth despite the smooth aggregate fundamentals. A number of studies have adopted this basic idea to analyze several important aspects of asset price dynamics, which I discuss in this section.

These models typically specify multiple groups of agents with heterogeneous beliefs about certain fundamental variables. It is well known that deriving dynamic equilibrium models with heterogeneous agents is a major challenge. Williams (1977), Detemple and Murthy (1994), Zapatero (1998), Basak (2000, 2005), and Jouini and Napp (2007) resolved this challenge by providing analytical solutions of dynamic asset market equilibrium in various continuous-time settings with complete markets.[5] In particular, Detemple and Murthy (1994) derived a production economy with logarithmic agents having heterogeneous beliefs about the unobservable growth of the production process. They used the equivalent-martingale approach developed by Cox and Huang (1986) to solve the equilibrium and showed that the equilibrium price of an asset is a wealth-weighted average of the prices that would prevail in economies populated by homogeneous agents with the beliefs of the respective agents in the original heterogeneous model. Zapatero (1998) analyzed a pure-exchange economy with logarithmic agents having heterogeneous beliefs about the aggregate endowment process. Basak (2000, 2005) analyzed a similar pure-exchange economy with agents having more general utility functions and heterogeneous beliefs about both fundamental and non-fundamental variables. It is well known that there exists a representative agent in a complete-market equilibrium. Basak used the representative-agent approach with stochastic weights for the respective agents to derive the equilibrium. Jouini and Napp (2007) constructed the so called "consensus belief" of the representative agent from the market equilibrium with agents holding heterogeneous beliefs to replicate the equilibrium dynamics under the assumption that the representative agent has the same preference as the individual agents. They showed that when agents have hyperbolic absolute risk aversion utility functions, the consensus belief implied by the equilibrium asset prices is a weighted average of the original individual beliefs with the weights giving by the individuals' risk tolerances (which are directly related to their wealth and consumption). Together, these models provided the basic framework for the later studies to analyze the effects of heterogeneous beliefs

---

[5] See also Gallmeyer and Hollifield (2008) for a dynamic equilibrium model with heterogeneous beliefs in an incomplete-market setting.

on asset price dynamics. Shefrin (2008) provided an eloquent treatment of some key mechanisms of dynamic asset market equilibrium with heterogeneous beliefs using discrete-time settings.

Endogenous Risk and Time-Varying Risk Premium

Incorporating heterogeneous beliefs leads to three important insights. First, the agents' endogenous wealth fluctuations can lead to excessive volatility relative to the level that would prevail in the absence of heterogeneous beliefs. The intuition works as follows. In the heterogeneous beliefs equilibrium, asset prices aggregate individual beliefs weighted by their risk tolerances. As optimists tend to take more aggressive positions, a positive shock causes them to gain greater wealth and larger risk tolerances relative to pessimists, which in turn give the optimists greater weights in determining the asset prices and thus push the prices even higher. Similarly, a negative shock reduces their wealth and risk tolerances, which in turn further suppress the asset prices. Thus, the agents' endogenous wealth acts as an amplification mechanism for the initial shock. This insight can help explain the so-called excess volatility observed in different asset returns, such as stock returns (e.g., Kurz (1996), Dumas, Kurshev and Uppal (2009) and Li (2007)) and bond yields (e.g., Xiong and Yan (2010) and Croitoru and Lu (2009)).

Second, the endogenous wealth fluctuations also cause the equilibrium risk premium to vary over time. This is because even if agents' equal weighted average beliefs are unbiased (i.e., equal to the objective probability measure that drives the economic uncertainty), the fluctuations of the relative wealth of optimists and pessimists can cause their risk-tolerance weighted beliefs to change over time. When the optimists' wealth is higher than the pessimists', the equilibrium asset prices may appear "expensive" from the viewpoint of an econometrician who holds the objective probability measure---i.e., the expected returns and risk premium are low. On the other hand, after a series of negative shocks, optimists will lose more wealth than pessimists as they tend to take more aggressive positions. The risk-tolerance weighted beliefs fall and suppress the asset prices. The expected returns and risk premium would then appear high from the view point of the econometrician. Xiong and Yan (2010) illustrated this mechanism by using a belief structure in which the dispersion of agents' beliefs changes over time although their average belief always tracks the objective belief of the econometrician. Their analysis shows that the time-varying risk premium induced by heterogeneous beliefs can help explain the failure of the expectations hypothesis in the bond markets and the ability of a tent-shaped linear combination of forward rates to predict bond returns (e.g., Cochrane and Piazzesi (2005)).

Third, the endogenous wealth fluctuations can also help explain the equity premium puzzle. As heterogeneous beliefs cause optimists and pessimists to trade against each other, the risk they face is not just the aggregate fundamental shocks but also includes the endogenous risk generated by the speculation

34

between them. David (2008) provided a model to calibrate the effects of such endogenous risk on the equilibrium equity premium. A lower risk aversion leads agents to speculate more against each other, and, as a result, each agent faces greater risk, even though the market price of risk decreases. When agents' coefficient of relative risk aversion is less than one, the increase in the amount of risk dominates the decrease in the market price of risk, generating a higher equity premium on net. David's calibration shows that endogenous risk causes half the observed equity premium and lowers the riskless interest rate by about 1%.

Note an important difference in affecting asset market dynamics between the heterogeneity in agents' beliefs and the heterogeneity in agents' endowments. In an influential paper, Constantinides and Duffie (1996) showed that heterogeneity in the form of uninsurable, persistent and heteroscedastic labor income shocks can help resolve the empirical difficulties encountered by representative-agent models through its ability to generate volatile consumption by individuals in an incomplete market setting. The heterogeneous labor incomes do not play a role in equilibrium if agents can efficiently share their labor income risks in complete markets. In contrast, heterogeneous beliefs can lead to volatile consumption by individuals even in complete market settings. This is because heterogeneous beliefs lead to speculation between agents regardless of whether the markets are complete or incomplete. In fact, as I discuss later, by completing the markets financial innovations can lead to more volatile, rather than smoother, consumption by individuals.

Shefrin (2008) advocated using state price density as a device to measure the overall sentiment of the market, as it aggregates the optimists and pessimists' belief distributions about future states based on their relative wealth. He provided numerical examples to show that when the belief distributions of both optimists and pessimists are log-normal, aggregating their heterogeneous beliefs in equilibrium leads to a multimodal distribution with fatter tails and more than one local maximum in the middle. This shape is consistent with the shape of the pricing kernel discovered in the work of Ait-Sahalia and Lo (2000) and Rosenberg and Engle (2002). The chapter by Barone-Adesi et al. (2011) in this handbook introduces a technique based on the state price density measured from stock and option data to measure fluctuations of market sentiment. In other related work, Buraschi and Jiltsov (2006) and Beber, Buraschi, and Breedon (2010) provided empirical evidence that the volatility smile of S&P500 index options and currency options is positively correlated with measures of belief dispersion.

Not just for the stock and bond markets, heterogeneous beliefs are also relevant for analyzing other markets as well. Dumas, Lewis, and Osambela (2011) invoked heterogeneous beliefs between domestic and foreign residents to explain a set of pricing anomalies in international finance, ranging from the home

equity bias and the co-movement of returns and international capital flows to the dependence of firm returns on local and foreign factors. Dieckmann (2011) and Chen, Joslin, and Tran (2011) analyzed the effects of heterogeneous beliefs on the sharing of disaster risks. Dieckmann (2011) showed that in the presence of heterogeneous beliefs about the likelihood of rare events, completing the markets by providing explicit rare event insurance contracts can lead to a higher, rather than lower, risk premium because the contracts cause individuals to increase rather than decrease their exposures to rare event risk based on the same endogenous risk argument I discussed earlier. Chen, Joslin, and Tran (2011) showed that when investors disagree about the likelihood of jump risk, a small amount of optimists in the economy can significantly reduce the disaster risk premium, and that the disaster risk premium can dramatically increase after the optimists suffer loses after a disaster.

Heterogeneous beliefs can also induce non-neutral effects on aggregate investment. Sims (2008) provided a model with two periods to illustrate this effect. His model features agents holding heterogeneous beliefs about inflation. Those who believe high inflation is less likely will find nominal lending attractive, while those who believe high inflation is more likely will find nominal borrowing a cheap source of financing. Heterogeneous beliefs motivate inflation optimists to borrow in nominal terms from inflation pessimists. Furthermore, if the agents have rates of relative risk aversion less (or higher) than one, each type chooses to consume less (or more) than what would have been consumed in the absence of heterogeneous beliefs. The net saving in turn leads to over- (or under-) investment. This non-neutral effect of heterogeneous beliefs about monetary policy prompts attention from policy makers.

Natural Selection

In the presence of agents whose beliefs are less reliable in predicting the future than others, the issue of natural selection emerges. Friedman (1953) long advocated that agents with inferior beliefs would eventually lose their money to others with superior beliefs and thus diminish in the markets. While compelling, this natural selection argument is not a foregone conclusion. In the recent literature, there is still an ongoing debate. By using a partial equilibrium setting, De Long et al. (1991) showed that agents with inferior beliefs may survive in the long run. This is because agents with modestly optimistic beliefs will bear more risk than those agents who hold the correct beliefs and, as a result, earn greater risk premium and grow their wealth on average at a higher rate. Their analysis ignores the potential price impact of the optimistic agents. After endogenizing the asset prices, a host of studies showed that the survival of agents with inferior beliefs may depend on the growth rate of the aggregate economy and agents' savings decisions and risk preferences. Blume and Easley (1992) showed that inferior beliefs can be beneficial for survival if savings decisions are exogenously given. Sandroni (2000) and Blume and

Easley (2006) endogenized both investment and savings decisions in settings with general risk preferences but bounded aggregate endowment processes. In their settings, agents with superior beliefs always dominate the market in the long run. Kogan et al. (2006) allowed unbounded aggregate endowment in a setting without intermediate consumption and showed that agents with inferior beliefs can survive. Their analysis especially emphasized that even in some cases in which agents with inferior beliefs do not survive in the long run, they could still have a significant price impact. Yan (2008) analyzed a setting with unbounded aggregate endowment, intermediate consumption, and agents with heterogeneity not only in beliefs but also in utility functions and time discount rates. In the absence of heterogeneity in utility function and time discount rate, inferior beliefs hurt long-run survival. However, when agents differ in risk preferences, survival also depends on the impact of beliefs on agents' savings decisions. Yan constructed a survival index to summarize the effects of these different dimensions on survival. Kogan et al. (2011) generalized their earlier work to a setting with intermediate consumption and more general utility functions, and provided necessary and sufficient conditions for agents with inferior beliefs to survive and to affect prices in the long run. In particular, they highlighted that agents with inferior beliefs can maintain a nontrivial consumption share and affect prices in an economy with decreasing absolute risk aversion preferences.

Despite the subtlety in the conclusions drawn by these studies based on different model settings, numerical calibrations tend to agree that even when agents with inferior beliefs fail to survive in the long run, their diminish can take over 100 years (e.g., Yan (2008) and Dumas, Kurshev and Uppal (2009)). With the steady birth of new investors over time, the natural selection argument does not seem to post a big concern over the potentially important effects of heterogeneous beliefs on asset markets.

VII. Welfare Analysis with Distorted Beliefs

Heterogeneous beliefs lead agents to engage in speculative transactions against each other. The previous sections summarize the large body of existing literature that shows the rich implications of heterogeneous beliefs. Some of the consequences such as over-trading, over-investment, and excessive leveraging have potential policy implications and prompt close attention from policy makers. To discuss policy-related issues, it is useful to measure social welfare in the presence of heterogeneous beliefs. It is useful to differentiate heterogeneous beliefs originated from prior beliefs and belief distortions. As discussed in Section III, prior beliefs depend on people's background and experience and are part of their preferences. It is difficult to attribute speculation caused by heterogeneous prior beliefs as excessive. However, there is ample empirical evidence that a wide range of psychological biases, such as overconfidence, distort investors' beliefs and affect their investment decisions. In the presence of

distorted beliefs, it is necessary that a social planner uses the objective probability measure that drives the economic uncertainty to evaluate agents' welfare on their behalves. However, in most realistic problems, the objective probability measure that drives economic uncertainty is not observable. Whose beliefs should the social planner use to analyze social welfare? This is an important challenge that confronts policy makers.

The recent work of Brunnermeier, Simsek and Xiong (2012) proposed a welfare criterion for models with distorted beliefs. Instead of choosing whose beliefs are superior, their idea is to assume that the objective probability measure lies between the agents' beliefs. In other words, the objective probability measure is either the beliefs of one of the agents or a convex combination of the agents' beliefs. As a result, one only needs to evaluate social welfare based on beliefs spanned by convex combinations of the agents' beliefs. This space is sufficiently large and contains all reasonable probability measures based on the given environment. Even if the objective probability measure lies outside this space, the analysis is still valid as everyone in the economy misses the objective measure. Based on this consideration, Brunnermeier, Simsek and Xiong proposed the following belief-neutral welfare criterion.

A Belief-Neutral Welfare Criterion: If a social allocation x is inefficient under any reasonable probability measure (i.e., any convex combination of all agents' beliefs), then it is belief-neutral inefficient.

There are two different approaches to implement this welfare criterion, one based on a given social welfare function and the other through the notion of Pareto efficiency. As is well known, these two approaches are internally consistent in standard settings with homogeneous beliefs because any Pareto efficient social allocation corresponds an optimal allocation that maximizes the aggregate expected utilities of agents under a set of non-negative weights.

This belief-neutral welfare criterion is restrictive, as it requires consistent ranking of one allocation relative to others under a large set of belief measures spanned by the convex combinations of all agents' beliefs. This condition implies that the welfare ranking is necessarily incomplete. Despite the incompleteness, this welfare criterion is particularly useful in detecting negative-sum games driven by agents' heterogeneous beliefs. For example, all agents might agree that they face a negative sum game but they might still proceed, because each agent thinks that he can win at the expense of others. In a sense, this welfare criterion extends the "externality view" to settings with distorted beliefs.

Brunnermeier, Simsek and Xiong (2012) provided a set of examples to show that this criterion can give clear-cut welfare ranking in many of the models discussed in this chapter. To illustrate the idea, I

will discuss one of their examples here. Consider a simple setting with two agents in an economy, A and B, and a single period. Each agent is endowed with ½ dollars and lives from $t = 0$ to $t = 1$. There is neither aggregate nor idiosyncratic endowment risk. Suppose that each agent consumes at $t = 1$ and has an increasing and strictly concave utility function $u(c_i)$. The two agents hold heterogeneous beliefs about a random variable, say f, which can take two possible values, either H or L. One may interpret this random variable as sun spot, which is independent of agents' endowment risk. Suppose that agent A assigns probability of π^A to state H, while agent B assigns π^B. The difference in beliefs ($\pi^A \neq \pi^B$) causes the agents to take on speculative trades against each other. We allow them to trade a contract that pays 1 if f takes the value of H and 0 otherwise. One can directly prove that the two agents will trade a positive amount on this contract, say k^A contracts by agent A and thus $-k^A$ by agent B, at a price of p. The exact values of k^A and p are not essential. It is important to note that as a result of this trade, the agents' consumption is not their deterministic endowments and would instead vary across the two states. Table I gives the allocation of consumption between the two agents in the status quo without the trading and in the market equilibrium with the trading driven by their heterogeneous beliefs.

Table I: Allocation of Consumption in Status Quo and Market Equilibrium

	Status Quo		Equilibrium with Heterogeneous Beliefs	
	State H	State L	State H	State L
Agent A	0.5	0.5	$0.5 + k^A(1 - p)$	$0.5 - k^A p$
Agent B	0.5	0.5	$0.5 - k^A(1 - p)$	$0.5 + k^A p$

For simplicity, suppose that the social planner uses so-called utilitarian social welfare function, which is the equally weighted sum of all agents' utility:

$$W(u^A, u^B) = u^A + u^B.$$

To evaluate the expected social welfare, the social planner uses a convex combination of the two agent's beliefs with a weight λ to agent A's beliefs and a weight $1 - \lambda$ to agent B's. In this measure, the probability of state H is $\pi = \lambda \pi^A + (1 - \lambda)\pi^B$. Then, the expected social welfare under the market equilibrium is

$$\pi\left[u\left(0.5 + k^A(1 - p)\right) + u\left(0.5 - k^A(1 - p)\right)\right] + (1 - \pi)[u(0.5 - k^A p) + u(0.5 + k^A p)].$$

The first square bracket above contains the social welfare in state H, while the second square bracket contains that in state L. By the strict concavity of the utility function $u(\cdot)$, the social welfare in each state

is less than $2u(0.5)$, the social welfare under the status quo. It is then straightforward to see that the expected social welfare under the market equilibrium is less than that under the status quo regardless of the particular convex combination of the agents' beliefs the social planner uses to evaluate the expectation.

In this example, the agents' speculative trading makes their consumption more volatile than their endowments, and thus reduces the sum of their expected utilities. That is, the agents face a negative-sum game in expected utility terms.[6] While this example is overly simplified, this negative-sum feature is common in all of the dynamic equilibrium models summarized in Section VI. One can adopt one of the settings in those models, with either continuous time or many discrete periods. If two agents have an identical utility function and equally shared endowments, it should be clear that trading driven by their heterogeneous beliefs about any random variable (either fundamental or non-fundamental) will make their consumption more volatile relative to the status quo and thus reduce the social welfare regardless of which beliefs we use to evaluate the social welfare.

Heterogeneous beliefs also induce negative-sum games in other types of models. In the resale option models of bubbles discussed in Section IV (e.g., Harrison and Kreps (1978), Morris (1996), Scheinkman and Xiong (2003), and Hong, Scheinkman and Xiong (2006)), agents are risk-neutral and any trading is a zero-sum game. Adding realistic trading costs make trading a negative-sum game. Still, agents trade and over-value an asset because they believe they can off-load the asset at an excessively high price to whomever will be the optimistic trader at the time. In the over-investment models discussed in Section IV (e.g., Bolton, Scheinkman and Xiong (2006), Gilchrist, Himmelberg, and Huberman (2005), and Panageas (2006)), the bubble component in asset prices induces each agent to prefer an investment scale that maximizes the firm's current market value at the expense of its long-run fundamentals, which determine the social welfare of risk-neutral agents. In the credit cycle models discussed in Section VI (e.g., Geanakoplos (2003, 2010), Fostel and Geanakoplos (2008), Simsek (2010), and He and Xiong (2012)), optimism induces optimists to take excessive leverage from pessimists to finance their initial investment, which exposes them to possible bankruptcy risk in the future. The presence of realistic bankruptcy costs again makes such leverage cycles a negative-sum game. In all of these settings, the welfare criterion proposed by Brunnermeier, Simsek and Xiong (2012) can generate clear-cut welfare ranking despite its incompleteness.

---

[6] The two agents may have unequal weights. Then, the speculative trading can act as a way to transfer wealth from the agent with smaller weight to the other with greater weight. However, the social planner can directly transfer wealth between them without inducing the increased consumption volatility. Following this logic, Brunnermeier, Simsek and Xiong (2012) provided detailed analysis to show that the speculative trading is also belief-neutral Pareto inefficient.

VIII. Summary and Future Directions

This chapter reviews the quickly growing body of work that builds on heterogeneous beliefs to explain anomalous asset market phenomena from bubbles, crises, and credit cycles to excess asset price volatility. Let me conclude with the two central threads of this literature. One is that in the presence of limits of arbitrage, heterogeneous beliefs lead to speculative investor behavior and asset price bubbles. This is because an asset owner values the option to resell the asset to a future optimist at a speculative profit. The other thread is that even in complete market settings with effective arbitrage, heterogeneous beliefs can nevertheless induce individuals to speculate against each other and lead to endogenous wealth fluctuations and endogenous price fluctuations.

These insights can help financial regulators and market participants understand and monitor risk (and especially systemic risk) in the financial system. In the aftermath of the recent financial crisis, which, according to many pundits, was ultimately driven by the two bubbles before the crisis---the Internet bubble and the housing bubble, it becomes imperative to develop an effective detection system for future bubbles. Such a system, if available, would allow financial regulators to monitor and control systemic risk in the financial system and help financial institutions avoid risk from the eventual collapses of bubbles. The heterogeneous beliefs based theories highlight several salient characteristics of bubbles, such as the joint occurrence of bubbles and frenzied trading, excessive leverages used by a set of market participants, and growing presence of inexperienced investors.   These characteristics should help developing such a detection system. However, it still remains an empirical challenge to systematically test the out-of-sample ability of these characteristics to predict future bubbles.

Another important empirical challenge is to identify the causes of financial institutions' risk takings during the market boom that preceded the recent crisis. With the help of hindsight, many commentators agreed that the "excessive" risks taken by financial institutions were a key factor leading to the ruining of their balance sheets during the crisis. However, the causes of their risk takings are still far from clear. They could be motivated by distorted incentives or distorted beliefs.   Both reasons can lead to unwarranted risk takings, but prompt very different policy responses. If the risk takings were driven by distorted incentives, appropriate regulations of bankers and traders' compensation can help prevent future financial crises. However, if the risk takings were caused by traders' disagreements about economic fundamentals, regulating traders' compensation (which is the focus of the recent Todd-Frank Act) is unlikely to resolve the key problem. Instead, improving information transparency and directly regulating risk takings might be more effective. The recent study of Cheng, Raina and Xiong (2012) designed a strategy to separate distorted beliefs from distorted incentives by examining personal home transactions

by a sample of financial industry employees and a matched sample of lawyers during the recent housing bubble. As home transactions are immune from the financial industry employees' job incentives, whether they were able to time the bubble or exercised cautions provide a lens to identify their beliefs about the housing market at the time. Interestingly, Cheng, Raina and Xiong found little evidence for financial industry employees either timing the bubble or exercising cautions, which suggests that they might have been too optimistic.

Many pundits have also pointed to innovative new securities as a vehicle for financial institutions to take/transfer risks during the recent market boom and bust. To evaluate risk in the financial system, it is important to have a thorough evaluation of the economic roles played by financial innovations. In an environment with heterogeneous beliefs, financial innovations can not only facilitate investors' risk sharing, but also allow them to engage in speculation against each other. Dieckman (2011) raised this issue in his analysis of rare event insurance contracts. Simsek (2011) presents an even more subtle effect that by helping investors to efficiently share their idiosyncratic risk, a financial innovation frees up their risk-bearing capacities to speculate more against each other based on their disagreements on another source of risk. As a result, the innovation leads the investors to have greater, rather than smaller, portfolio volatility. Shen, Yan, and Zhang (2011) propose an important mechanism that financial innovations facilitate speculation by reducing collateral requirements. Together, these papers highlight that the tradeoff between risk sharing and speculation can have important welfare implications and prompt close attention from financial regulators.

References

Abreu, Dilip and Markus K. Brunnermeier (2002), Synchronization risk and delayed arbitrage, *Journal of Financial Economics* 66, 341-360.

Abreu, Dilip and Markus K. Brunnermeier (2003), Bubbles and crashes, *Econometrica* 71, 173-204.

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2009), Fragility of asymptotic agreement under Bayesian learning, Working paper, MIT.

Ackert, Lucy, Narat Charupat, Bryan Church, and Richard Deaves (2006), Margin, short selling, and lotteries in experimental asset markets, *Southern Economic Journal* 73, 419–36.

Adrian, Tobias and Hyun Song Shin (2010), Liquidity and leverage, *Journal of Financial Intermediation* 19, 418-437.

Ait-Sahalia, Yacine and Andrew Lo (1998), Nonparametric estimation of state‐price densities implicit in financial asset prices, *Journal of Finance* 53, 499-547.

Allen, Franklin and Douglas Gale (2000), Bubbles and crises, *Economic Journal* 110, 236-55.

Allen, Franklin and Douglas Gale (2007), *Understanding Financial Crises*, Oxford University Press, New York.

Allen, Franklin and Gary Gorton (1993), Churning bubbles, *Review of Economic Studies* 60, 813-36.

Alpert, M. and H. Raiffa (1982), A progress report on the training of probability assessors, in Daniel Kahneman, Paul Slovic, and Amos Tversky, ed.: *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

Anderson, Evan, Eric Ghysels, and Jennifer Juergens (2005), Do heterogeneous beliefs matter for asset pricing? *Review of Financial Studies* 18, 875-924.

Aumann, Robert (1976), Agreeing to disagree, *Annals of Statistics* 4, 1236-1239.

Baker, Malcolm, Jeremy Stein, and Jeffrey Wurgler (2003), When does the market matter? Stock prices and the investment of equity-dependent firms, *Quarterly Journal of Economics* 118, 969–1006.

Banerjee, Snehal (2011), Learning from prices and the dispersion in beliefs, *Review of Financial Studies* 24, 3025-3068.

Barber, Brad and Terrance Odean (2000), Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.

Barber, Brad, Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean (2009), Just how much do individual investors lose by trading? *Review of Financial Studies* 22, 609-632.

Barberis, Nicholas (2011), Psychology and the financial crisis of 2007-2008, Working paper, Yale University.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998), A model of investor sentiment, *Journal of Financial Economics* 49, 307-343.

Barberis, Nicholas and Richard Thaler (2003), A survey of behavioral finance, In: Constantinides, G., Harris, M., Stulz, R. (Eds.), *Handbook of the Economics of Finance*, North-Holland, Amsterdam.

Basak, Suleyman (2000), A model of dynamic equilibrium asset pricing with heterogeneous beliefs and extraneous risk, *Journal of Economic Dynamics and Control* 24, 63-95.

Bebchuk, Lucian and Jesse Fried (2004), *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*, Harvard University Press, MA.

Ben David, Itzhak, John Graham, and Campbell Harvey (2010), Managerial Miscalibration, NBER working paper 16215.

Benabou, Roland (2009), Groupthink: Collective delusions in organizations and markets, Working paper, Princeton University.

Blanchard, Oliver, Changyong Rhee, and Lawrence Summers (1993), The stock market, profit and investment, *Quarterly Journal of Economics* 107, 115-136.

Blanchard, Olivier and Mark Watson (1983), Bubbles, rational expectations and financial markets, National Bureau of Economic Research, NBER Working Papers 945.

Blume, Lawrence and David Easley (1992), Evolution and market behavior, *Journal of Economic Theory* 58, 9-40.

Blume, Lawrence and David Easley (2006), If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets, *Econometrica* 74, 929-966.

Bolton, Patrick, Jose Scheinkman, and Tano Santos (2011), Outside and inside liquidity, *Quarterly Journal of Economics* 126, 259-321.

Bolton, Patrick, Jose Scheinkman, and Wei Xiong (2005), Pay for short-term performance: Executive compensation in speculative markets, *Journal of Corporation Law* 30, 721-747.

Bolton, Patrick, Jose Scheinkman, and Wei Xiong (2006), Executive compensation and short-termist behavior in speculative markets, *Review of Economic Studies* 73, 577-610.

Bris, Arturo, William Goetzmann, and Ning Zhu (2007), Efficiency and the bear: Short sales and markets around the world, *Journal of Finance* 62, 1029-1079.

Brunnermeier, Markus and Lasse Pedersen (2009), Funding liquidity and market liquidity, *Review of Financial Studies* 22, 2201-2238.

Brunnermeier, Markus, Alp Simsek and Wei Xiong (2012), A welfare criterion for models with distorted beliefs, Working paper, Princeton University.

Buraschi, Andrea and Alexei Jiltsov (2006), Model uncertainty and option markets with heterogeneous beliefs, *Journal of Finance* 61, 2841-97.

Campbell, John, Andrew Lo, and Craig MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, NJ.

Cao, Dan (2011), Collateral shortages, asset price and investment volatility with heterogeneous beliefs, Working paper, Georgetown University.

Chen, Hui, Scott Joslin, and Ngoc-Khanh Tran (2011), Rare disasters and risk sharing with heterogeneous beliefs, Working paper, MIT.

Chen, Joseph, Harrison Hong, and Jeremy Stein (2002), Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171-205.

Cheng, Ing-haw, Sahil Raina, and Wei Xiong (2012), Wall Street and the housing bubble: Bad incentives, bad models, or bad luck?, Working paper, Princeton University.

Chinco, Alex and Chris Mayer, Noise traders, distant speculators and asset bubbles in the housing market, Working paper, Columbia University.

Constantinides, George and Darrell Duffie (1996), Asset pricing with heterogeneous consumers, *Journal of Financial Economics* 104, 219-240.

Coval, Joshua and Erik Stafford (2007), Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479-512.

Croitoru, Benjamin and Lei Lu (2009), Asset pricing in a monetary economy with heterogeneous beliefs, Working paper, McGill University.

Danielsson, Jon, Jean-Pierre Zigrand, and Hyun Song Shin (2010), Balance sheet capacity and endogenous risk, Working paper, London School of Economics and Princeton University.

Daniel, Kent, David Hirshleifer, and A. Subrahmanyam (1998), Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839-1885.

David, Alexander (2008), Heterogeneous beliefs, speculation, and the equity premium, *Journal of Finance* 63, 41-83.

D'Avolio, Gene (2002), The market for borrowing stock, *Journal of Financial Economics* 66, 271-306.

Diether, Karl, Christopher Malloy, and Anna Scherbina (2002), Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113-2141.

De Long, Brad, Andrei Shleifer, Lawrence Summers, and Robert Waldman (1990), Noise trader risk in financial market, *Journal of Political Economy* 98, 703-738.

De Long, Brad, Andrei Shleifer, Lawrence Summers, and Robert Waldman (1991), The survival of noise traders in financial markets, *Journal of Business* 64, 1-19.

Detemple, Jerome and Shashidhar Murthy (1994), Intertemporal asset pricing with heterogeneous beliefs, *Journal of Economic Theory* 62, 294-320.

Dieckmann, Stephan (2011), Rare event risk and heterogeneous beliefs: The case of incomplete markets, *Journal of Financial and Quantitative Analysis* 46, 459-488.

Duffie, Darrel (2010), Asset price dynamics with slow-moving capital, (American Finance Association Presidential Address), *Journal of Finance* forthcoming.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen (2002), Securities lending, shorting, and pricing, *Journal of Financial Economics* 66, 307-339.

Duffie, Darrel, Nicolae Garleanu, and Lasse Heje Pedersen (2005), Over-the-counter markets, *Econometrica* 73, 1815-1847.

Dufwenberg, Martin, Tobias Lindqvist, and Evan Moore (2005), Bubbles and experience: An experiment on speculation, *American Economic Review* 95, 1731–1737.

Dumas, Bernard, Alexander Kurshev, and Raman Uppal (2009), Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility, *Journal of Finance* 64, 579-629.

Dumas, Bernard, Karen Lewis, and Emilio Osambela (2011), Differences of opinion and international equity markets, Working paper, Wharton School.

Fostel, Ana and John Geanakoplos (2008), Leverage cycles and the anxious economy, *American Economic Review* 98, 1211-1244.

Friedman, Milton (1953), *Essays in Positive Economics,* University of Chicago Press, Chicago, IL.

Froot, Kenneth and Paul O'Connell (1999), The pricing of US catastrophe reinsurance, in Kenneth Froot ed. *The Financing of Catastrophe Risk*, University of Chicago Press.

Galbraith, John Kenneth (1997), *The Great Crash 1929*, Houghton Mifflin Company, New York.

Gallmeyer Michael and Burton Hollifield (2008), An examination of heterogeneous beliefs with a short-sale constraint in a dynamic economy, *Review of Finance* 12, 323-364.

Garber, Peter (2000), *Famous First Bubbles: The Fundamentals of Early Manias*, MIT Press, Cambridge, Mass.

Garleanu, Nicolae and Lasse Pedersen (2011), Margin-based asset pricing and deviations from the law of one price, *Review of Financial Studies* 24, 1980-2022.

Geanakoplos, John (2003), Liquidity, default, and crashes: Endogenous contracts in general equilibrium, In *Advances in Economics and Econometrics Theory and Applications, Eighth World Conference,* Vol 2, 170-205, Econometric Society Monographs.

Geanakoplos, John (2010), The leverage cycle, *NBER Macroeconomics Annual 2009,* 24, 1-65.

Geczy, Christopher, David Musto, and Adam Reed (2002), Stocks are special too: an analysis of the equity lending market, *Journal of Financial Economics* 66, 241-269.

Gervais, Simon, and Terrance Odean (2001), Learning to be overconfident, *Review of Financial Studies* 14, 1-27.

Gilchrist, Simon, Charles Himmelberg, and Gur Huberman (2005), Do stock price bubbles influence corporate investment? *Journal of Monetary Economics* 52, 805-827.

Grinblatt, Mark and Matti Keloharju (2000), The investment behavior and performance of various investor types: a study of Finland's unique data set, *Journal of Finance* 55, 43-67.

Griffin, D. and A. Tversky (1992), The weighing of evidence and the determinants of overconfidence, *Cognitive Psychology* 24, 411–435.

Gromb, Denis and Dimitri Vayanos (2002), Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361-407.

Gromb, Denis and Dimitri Vayanos (2010), Limits of Arbitrage: The state of the theory, *Annual Review of Financial Economics* 2, 251-275.

Grossman, Sanford and Joseph Stiglitz (1980), On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Harris, Milton and Artur Raviv (1993), Differences of opinion make a horse race, *Review of Financial Studies* 6, 473-506

Harrison, Michael, and David M. Kreps (1978), Speculative investor behavior in a stock-market with heterogeneous expectations, *Quarterly Journal of Economics* 92, 323-336.

Haruvy, Ernan and Charles Noussair (2006), The effect of short selling on bubbles and crashes in experimental spot asset markets, *Journal of Finance* 61, 1119-1157.

Haruvy, Ernan, Yaron Lahav, and Charles Noussair (2007), Traders' expectations in asset markets: experimental evidence, *American Economic Review* 97, 1901-1920.

Haughwout, Andrew, Donghoon Lee, Joseph Tracy, and Wilbert van der Klaauw (2011), Real estate investors, the leverage cycle and the housing market crisis, Working paper, Federal Reserve Bank of New York.

He Zhiguo and Arvind Krishnamurthy (2009), Intermediary asset pricing, Working paper, University of Chicago and Northwestern University.

He, Zhiguo and Wei Xiong (2010), Delegated asset management and investment mandates, Working paper, University of Chicago and Princeton University.

He, Zhiguo and Wei Xiong (2012), Debt financing in asset markets, *American Economic Review Papers and Proceedings*, forthcoming.

Hirota, Shinichi, and Shyam Sunder (2007), Price bubbles sans dividend anchors: Evidence from laboratory stock markets, *Journal of Economic Dynamics and Control* 31, 1875-1909.

Hirshleifer, David (2001), Investor psychology and asset pricing, *Journal of Finance* 56, 1533-1597.

Hong, Harrison and Jeffrey Kubik (2003), Analyzing the analysts: Career concerns and biased earnings forecasts, *Journal of Finance* 58, 313-351.

Hong, Harrison, Jose Scheinkman, and Wei Xiong (2006), Asset float and speculative bubbles, *Journal of Finance* 61, 1073-1117.

Hong, Harrison and Jeremy Stein (2003), Differences of opinion, short‐sales constraints, and market crashes, *Review of Financial Studies* 16, 487-525.

Hong, Harrison and Jeremy Stein (2007), Disagreement and the stock market, *Journal of Economic Perspectives* 21, 109-128.

Hong, Harrison and David Sraer (2011), Quiet bubbles, Working paper, Princeton University.

Hussam, Reshmaan, David Porter, and Vernon Smith (2008), Thar she blows: Can bubbles be rekindled with experienced subjects? *American Economic Review* 98, 924-937.

Jarrow, Robert (1980), Heterogeneous expectations, restrictions on short-sales, and equilibrium asset prices, *Journal of Finance* 35, 1105-1113.

Jouini, Elyes and Clotilde Napp (2007), Consensus consumer and intertemporal asset pricing with heterogeneous beliefs, *Review of Economic Studies* 74, 1149-1174.

Kandel, Eugene and Neil Pearson (1995), Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, 831-872.

Kindleberger, Charles (1978), *Manias, Panics, and Crashes: A History of Financial Crises*, John Wiley & Sons, New York.

Kiyotaki, Nobu and John Moore (1997), Credit cycles, *Journal of Political Economy* 105, 211-248.

Kondor, Peter (2009), Risk in dynamic arbitrage: The price effects of convergence trading, *Journal of Finance* 64, 631-655.

Kogan, Leonid, Stephen Ross, Jiang Wang, and Mark Westerfield (2006), The price impact and survival of irrational traders, *Journal of Finance* 61, 195-229.

Kogan, Leonid, Stephen Ross, Jiang Wang, and Mark Westerfield (2011), Market selection, Working paper, MIT.

Krishnamurthy, Arvind (2009), Amplification mechanisms in liquidity crises, *American Economic Journals-Macroeconomics* 2, 1-33.

Kurz, Mordecai (1994), On the Structure and Diversity of Rational Beliefs, *Economic Theory* 4, 877-900.

Kurz, Mordecai (1996), Rational beliefs and endogenous uncertainty, *Economic Theory* 8, 383-397.

Kyle, Albert (1985), Continuous auctions and insider trading, *Econometrica* 53, 1315-1335.

Kyle, Albert and Albert Wang (1997), Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *Journal of Finance* 52, 2073–90.

Kyle, Albert and Wei Xiong (2001), Contagion as a wealth effect, *Journal of Finance* 56, 1401-1440.

Lamont, Own and Richard Thaler (2003), Can the market add and subtract? Mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227-268.

Lei, Vivian, Charles Noussair, and Charles Plott (2001), Non-speculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality, *Econometrica* 69, 830–859.

Lin, H. and M. McNichols (1998), Underwriting relationships, analysts' earnings forecasts and investment recommendations, *Journal of Accounting and Economics* 25, 101-127.

Mackay, Charles (1841), *Extraordinary Popular Delusions and the Madness of Crowds*, London.

Malmendier, Ulrike and D. Shantikumar (2007), Are investors naive about incentives? *Journal of Financial Economics* 85, 457-489.

Mankiw, Gregory, Ricardo Reis, and Justin Wolfers (2004), Disagreement about inflation expectations, *NBER Macroeconomics Annual 2003*, 18, 209-248.

Mei, Jianping, Jose Scheinkman, and Wei Xiong (2009), Speculative trading and stock prices: Evidence from Chinese A-B share premia, *Annals of Economics and Finance* 10, 225-255.

Mian, Atif and Amir Sufi (2009), The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis, *Quarterly Journal of Economics* 124, 1449-1496.

Milgrom, Paul and Nancy Stokey (1982), Information, trade and common knowledge, *Journal of Economic Theory* 26, 17-27.

Miller, Edward (1977), Risk, uncertainty and divergence of opinion, *Journal of Finance* 32, 1151-1168.

Minsky, Hyman (1974), The modeling of financial instability: An introduction, *Proceedings of the Fifth Annual Pittsburgh Conference.*

Mitchell, Mark, Lasse Heje Pedersen and Todd Pulvino (2007), Slow moving capital, *American Economic Review* 97, 215-220.

Morck, Randall, Andrei Shleifer and Robert Vishny (1990), The stock market and investment: Is the market a sideshow?, *Brookings Papers on Economic Activity* 2, 157-215.

Morris, Stephen (1995), The common prior assumption in economic theory, *Economics and Philosophy* 11, 227-253.

Morris, Stephen (1996), Speculative investor behavior and learning, *Quarterly Journal of Economics* 111, 1111-1133.

Odean, Terrance (1998), Volume, volatility, price, and profit when all traders are above average, *Journal of Finance* 53, 1887-1934.

Odean, Terrance (1999), Do investors trade too much?, *American Economic Review* 89, 1279-1298.

Ofek, Eli and Matthew Richardson (2003), DotCom Mania: The rise and fall of Internet stock prices, *Journal of Finance* 58, 1113-1137.

Panageas, Stavros (2006), The neoclassical q theory of investment in speculative markets, Working paper, University of Chicago.

Pastor, Lubos, and Pietro Veronesi (2006), Was there a nasdaq bubble in the late 1990s? *Journal of Financial Economics* 81, 61-100.

Pastor, Lubos, and Pietro Veronesi (2009), Learning in financial markets, *Annual Review of Financial Economics* 1, 361-381.

Polk, Christopher and Paola Sapienza (2009), The stock market and corporate investment: A test of catering theory, *Review of Financial Studies* 22, 187-217.

Porter, David, and Vernon Smith (1995), Futures contracting and dividend uncertainty in experimental asset markets, *Journal of Business* 68, 509-541.

Rajan, Raghuram (2010), *Fault Lines*, Princeton University Press.

Reinhart, Carmen and Kenneth Rogoff (2009), *This Time Is Different*, Princeton University Press, NJ.

Rosenberg, Joshua and Robert Engle (2002), Empirical pricing kernels, *Journal of Financial Economics* 64, 341-372.

Rothschild, Michael (1974), A two-armed bandit theory of market pricing, *Journal of Economic Theory* 9, 185-202.

Sandroni, Alvaro (2000), Do markets favor agents able to make accurate predictions? *Econometrica* 68, 1303-1342.

Savage, L. J. (1954), *The Foundation of Statistics*, John Wiley and Sons.

Scheinkman, Jose A., and Wei Xiong (2003), Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183-1219.

Sebenius, James and John Geanakoplos (1983), Don't bet on it: Contingent agreements with asymmetric information, *Journal of the American Statistical Association* 78, 424-426.

Shefrin, Hersh (2008), *A Behavioral Approach to Asset Pricing*, Academic Press.

Shen, Ji, Hongjun Yan, and Jinfan Zhang (2011), Collateral-motivated financial innovation, Working paper, Yale University.

Shiller, Robert (1998), *Macro Markets*, Oxford University Press, NY.

Shiller, Robert (2000), *Irrational Exuberance*, Princeton University Press, Princeton, NJ.

Shleifer, Andrei and Robert Vishny (1997), Limits of arbitrage, *Journal of Finance* 52, 35-55.

Sims, Christopher (2008), Inflation expectations, uncertainty, and monetary policy, Working paper, Princeton University.

Simsek, Alp (2010), Belief disagreements and collateral constraints, Working paper, Harvard University.

Simsek, Alp (2011), Speculation and risk sharing with new financial assets, Working paper, Harvard University.

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams (1988), Bubbles, crashes, and endogenous expectations in experimental spot asset markets, *Econometrica* 56, 1119-1151.

Stein, Jeremy (1996), Rational capital budgeting in an irrational world, *Journal of Business* 69, 429-455.

Welch, Ivo (2000), Views of financial economists on the equity premium and on professional controversies, *Journal of Business* 73, 501-537.

White, Eugene (1990), The stock market boom and crash of 1929 revisited, *Journal of Economic Perspectives* 4, 67-83.

Xiong, Wei (2001), Convergence trading with wealth effects: An amplification mechanism in financial markets, *Journal of Financial Economics* 62, 2001, 247-292.

Xiong, Wei and Hongjun Yan (2010), Heterogeneous expectations and bond markets, *Review of Financial Studies* 23, 1433-1466.

Xiong, Wei and Jialin Yu (2011), The Chinese warrants bubble, *American Economic Review* 101, 2723-2753.

Yan, Hongjun (2008), Natural selection in financial markets: Does it work? *Management Science* 54, 1935-1950.

Yu, Jialin (2011), Disagreement and return predictability of stock portfolios, *Journal of Financial Economics* 99, 162-183.

Zapatero, Fernando (1998), Effects of financial innovations on market volatility when beliefs are heterogeneous, *Journal of Economic Dynamics and Control* 22, 597-626.