Market Ecology and the Economics of Crisis*

Rajiv Sethi†

January 12, 2012

Introduction

In his autobiographical essay *The Ins and Outs of Late Twentieth Century Economics*, Duncan Foley observed that "economics in the late 1960s suffered from a classical scientific dilemma in that it had two theories, the microeconomic general equilibrium theory, and the macroeconomic Keynesian theory, each of which seemed to have considerable explanatory power in its own domain, but which were incompatible." Specifically:

> The general equilibrium theory forged by Walras and elaborated by Wald (1951), McKenzie (1959), and Arrow and Debreu (1954) can be used, with the assumption that markets exist for all commodities at all future moments and in all contingencies, to represent macroeconomic reality by simple aggregation. The resulting picture of macroeconomic reality, however, has several disturbing features. For one thing, competitive general equilibrium is efficient, so that it is incompatible with the unemployment of any resources productive enough to pay their costs of utilization. This is difficult to reconcile with the common observation of widely fluctuating rates of unemployment of labor and of capacity utilization of plant and equipment. General equilibrium theory reduces economic production and exchange to the pursuit of directly consumable goods and services, and as a result has no real role for money... The general equilibrium theory

*An essay in honor of Duncan Foley. I'm grateful to Alan Kirman and Dick Nelson for comments on an earlier version, and to Duncan for guidance and support throughout my career.

†Department of Economics, Barnard College, Columbia University and the Santa Fe Institute.

> can accommodate fluctuations in output and consumption, but only as responses to external shocks to resource availability, technology or tastes. It is difficult to reconcile these relatively slowly moving factors with the large business-cycle fluctuations characteristic of developed capitalist economies. In assuming the clearing of markets for all contingencies in all periods, general equilibrium theory assures the consistency... of individual consumption, investment, and production plans, which is difficult to reconcile with the recurring phenomena of financial crisis and asset revaluation that play so large a role in actual capitalist economic life.

Foley contrasted this with Keynes' theory, which appeared to conform better to economic reality but suffered from problems of its own:

> Keynes views money as central to the actual operation of developed capitalist economies, precisely because markets for all periods and contingencies do not exist to reconcile differences in agents' opinions about the future. Because agents cannot sell all their prospects on contingent claims markets, they are liquidity constrained. In a liquidity constrained economy there is no guarantee that all factor markets will clear without unemployed labor or unutilized productive capacity. Market prices are inevitably established in part by speculation on an uncertain future. As a result the economy is vulnerable to endogenous fluctuations as the result of herd psychology and self-fulfilling prophecy. From this point of view it is not hard to see why business cycle fluctuations are a characteristic of a productively and financially developed capitalist economy, nor why the potential for financial crisis is inherent in decentralized market allocation of investment... But there are many loose ends in Keynes' argument. In presenting the equilibrium of short-term expectations that determines the level of output, income and employment in the short period, for example, Keynes argues that entrepreneurs hire labor and buy raw materials to undertake production because they form an expectation as to the volume of sales they will achieve when the production process runs its course... But Keynes offers no systematic alternative account of how entrepreneurs form a view of their prospects on the market to take the place of the assumption of perfect competition and market clearing. This turns out, in detail, to be a very difficult problem to solve.

Foley goes on to describe his pursuit, in collaboration with his MIT colleague Miguel Sidrauski, of the "project of a macroeconomic theory distinct from Walrasian general equilibrium theory." They did so by building on the Hicks' (1946) notion of *temporary equilibrium*, which allows for the clearing of spot markets but without requiring that individual plans be mutually consistent. Their two-sector model incorporated markets for stocks of money, bonds, and capital assets, as well as flows of labor, investment and consumption goods. They did not, however, provide an account of expectation formation, and therefore left unfilled one of the major gaps in the *General Theory*.

In the meantime, the economics profession was moving in an entirely different direction, by adopting the methodological imperative that the expectations of individuals within any model be consistent with the model itself. Foley recognized earlier than most that this so-called "rational expectations" hypothesis led right back to the general equilibrium model that he had earlier rejected:

> In my view, the rational expectations assumption which Lucas and Sargent put forward to "close" the Keynesian model, was only a disguised form of the assumption of the existence of complete futures and contingencies markets. When one unpacked the "expectations" language of the rational expectations literature, it turned out that these models assumed that agents formed expectations of futures and contingency prices that were consistent with the aggregate plans being made, and hence were in fact competitive general equilibrium prices in a model of complete futures and contingency markets... What the profession took to be an exciting breakthrough in economic theory I saw as a boring and predictable retracing of an already discredited path.

This passage describes very clearly what is often left implicit and unrecognized in contemporary macroeconomics: the rational expectations hypothesis entails the mutual consistency of individual plans without providing for a mechanism through which such consistency could be attained. In the case of complete markets, any inconsistency of plans is revealed immediately in the form of a disequilibrium in some futures market. But in the absence of complete markets, it is entirely possible for such inconsistency to remain concealed, even while all spot markets (and those futures markets that exist) clear. Only with the passage of time can this inconsistency be revealed, as expectations are *discovered* to have been incorrect. Such discoveries induce belief revision at the level of individual households and firms. The rational expectations hypothesis presumes, in effect, that this dynamic adjustment process quickly and reliably corrects any mutual inconsistencies in plans that might arise. Viewed in this manner, the hypothesis of rational expectations is an equilibrium assumption, usually made without attention to any equilibrating process.

The idea that this modeling hypothesis has severe limitations has become increasingly clear in the wake of the recent financial crisis. Woodford (2011) argues, for instance, that the "macroeconomics of the future... will have to go beyond conventional late-twentieth-century methodology... by making the formation and revision of expectations an object of analysis in its own right, rather than treating this as something that should already be uniquely determined once the other elements of an economic model (specifications of preferences, technology, market structure, and government policies) have been settled."

Hence the problem of expectations in macroeconomics and finance, left unresolved by Keynes and later by Foley and Sidrauski, remains open. In this essay I argue that progress on this question requires that one allow for *heterogeneity* across individuals in beliefs and forecasting rules, and explicit consideration of the *process* by which information is absorbed by asset prices through the activity of traders. The distribution of forecasting rules affects the dynamics of asset prices and hence the relative profitability of the various strategies in use. This in turn causes changes in the distribution of strategies over time. I argue that this *ecological* approach to markets can provide us with insights into a broad range of phenomena, including bubbles and crashes in speculative asset markets as well as large procyclical movements in leverage and maturity transformation within the non-financial sector. I begin with the example of event that briefly rocked financial markets not long ago, and that serves to illustrate the manner in which interacting strategies can give rise to complex asset price movements that have feedback effects on the evolving distribution of strategies over time.

The Flash Crash

On the 6th of May, 2010, an extraordinary event occurred in U.S. financial markets. Stock prices plunged to levels that wiped out close to a trillion dollars in value, and then promptly recovered to erase most of these losses. Some securities traded at prices that were absurdly inconsistent with any notion of their intrinsic worth. Shares in Accenture fell to a penny, while those in Sotheby's rose to a hundred thousand dollars apiece. These were not simply nominal quotes in an order book: volume remained heavy throughout the episode and a large number of transactions occurred at these prices. The exchanges subsequently cancelled all trades occurring at prices that departed by more than sixty percent from the previous close, reversing transactions in almost three hundred securities.

What cause this *flash crash*, as the event has come to be known? Kirilenko *et al.* (2011) attempted to answer this question by examining audit-trail data for all transactions in the June 2010 E-mini S&P 500 futures contract over the four day period May 3-6. What emerges from their analysis is a rich description of an ecology of trading strategies that interact to jointly determine the dynamics of prices. The authors classify accounts into six categories based on patterns exhibited in their trading behavior, such as horizon length, order size, and the willingness to accumulate significant net positions. The categories are High Frequency Traders (HFTs), Intermediaries, Fundamental Buyers, Fundamental Sellers, Opportunistic Traders and Small Traders:

> [Different] categories of traders occupy quite distinct, albeit overlapping, positions in the ecosystem of a liquid, fully electronic market. HFTs, while very small in number, account for a large share of total transactions and trading volume. Intermediaries leave a market footprint qualitatively similar, but smaller to that of HFTs. Opportunistic Traders at times act like Intermediaries (buying a selling around a given inventory target) and at other times act like Fundamental Traders (accumulating a directional position).

Based on this taxonomy, the authors examine the manner in which the strategies vary with respect to trading volume, liquidity provision, directional exposure, and profitability. Although high-frequency traders constitute a minuscule proportion (about one-tenth of one percent) of total accounts, they are responsible for more than a third of aggregate trading volume in this market. They have extremely short trading horizons and maintain low levels of directional exposure. Under normal market conditions they are net providers of liquidity but their desire to avoid significant exposure means that they can become liquidity takers very quickly and on a large scale. This appears to have happened during the crash:

> During the Flash Crash, the trading behavior of HFTs, appears to have exacerbated the downward move in prices. High Frequency Traders who initially bought contracts from Fundamental Sellers, proceeded to sell contracts and compete for liquidity with Fundamental Sellers. In addition, HFTs appeared to rapidly buy

> and [sell] contracts from one another many times, generating a hot potato effect before Opportunistic or Fundamental Buyers were attracted by the rapidly falling prices to step in and take these contracts off the market.

The authors find that high frequency traders in this market appear able to anticipate (and profit from) price movements over very short horizons:

> High Frequency Traders effectively predict and react to price changes... [they] are consistently profitable although they never accumulate a large net position... HFTs appear to trade in the same direction as the contemporaneous price and prices of the past five seconds. In other words, they buy... if the immediate prices are rising. However, after about ten seconds, they appear to reverse the direction of their trading... possibly due to their speed advantage or superior ability to predict price changes, HFTs are able to buy right as the prices are about to increase.

During the crash itself the fastest moving traders with the most effective algorithms for short run price prediction were able to trade ahead of their slower and less effective brethren, imposing significant losses on the latter. Since no trades were cancelled in this particular market, these losses involved a significant transfer of wealth in a short period of time.

This analysis of the crash illustrates several features of a speculative asset market. There is considerable heterogeneity in trading strategies, with a large proportion of trades being made on the basis of rapid responses to incoming market data rather than fundamental research. The dynamics of asset prices itself depends on this composition of trading strategies. And endogenous transfers of wealth, which can be significant during periods of instability, have the effect of altering the distribution of trading strategies in the market.

These features are not specific to the particular market considered here. At the time of the crash, more than sixty percent of transactions in U.S. equities involved orders placed by algorithms (Grant and MacKenzie, 2010). The strategies implemented by such programs are typically characterized by extremely short holding periods, amounting to no more than a few seconds on average, and very limited overnight directional exposure. More importantly, they neglect entirely any assessment of the stream of revenue to which ownership of an asset gives title. As such, they are the modern incarnation of technical trading strategies based on the identification of patterns in tables and charts. The speed, scale, and automation are new, but the strategies themselves have been around for as long as speculative asset markets have existed.

The flash crash was an extreme event with respect to the scale of departures of prices from any reasonable assessment of their intrinsic value, and the speed with which these departures arose and were corrected. But it was also routine in the sense that such departures do arise from time to time, building cumulatively rather than suddenly, and lasting for months or years rather than minutes, with corrections that can be rapid or prolonged but almost impossible to time. Understanding the crash is accordingly a step towards understanding such disruptions more generally.

Market Ecology

The number of distinct trading strategies in an asset market is as large as the number of active traders, since no two market participants adopt perfectly identical strategies. Nevertheless, it is analytically useful to partition the set of strategies into two categories: those based on an assessment of the intrinsic value of an asset, and those based on the use of market data to make short-term price forecasts. The terms fundamental and technical analysis broadly capture these two approaches to trading.

Fundamental analysis provides the channel through which information about the earnings flows to which the asset gives title (or the risk-sensitive rates at which these should be discounted) comes to be reflected in the asset's price. In a market dominated by such strategies, price changes will reflect changes over time in beliefs about fundamentals, as individual traders identify mispriced securities and trade on this information. As long as this process operates rapidly and smoothly, price volatility should be commensurate with the volatility in earnings and other measures of an asset's intrinsic worth.

But no speculative market is dominated by fundamental traders. (In the Kirilenko study of the S&P E-mini futures market, such strategies accounted for about 12 percent of trading volume.) When both technical and fundamental strategies are present, the asset price dynamics can be more complex and sensitive to the composition of strategies in the trading population. Beja and Goldman (1980) illustrated this point in a simple model with two trader types. Their main insight was that if the prevalence of momentum based strategies was too large relative to that of strategies based on fundamental analysis, then the dynamics of asset prices would be locally unstable: departures of prices from fundamentals would be amplified rather than corrected over time. More importantly, they argued that the relationship between the composition of strategies and market stability was discontinuous: there was a threshold value of this population mixture that separated the stable from the unstable regime, and an imperceptible change in composition that took the market across the threshold could result in dramatic increases in volatility.

A key feature of the Beja-Goldman model is that prices do not adjust instantaneously to clear the asset market, and this allows price movements to be informative and to exhibit some degree of short-run momentum. Technical strategies can exploit this momentum and accelerate the convergence of prices to fundamentals, provided that such strategies remain relatively rare. But if the use of technical analysis becomes too widespread in the market, the dynamics lose stability: prices start to overshoot fundamentals, resulting in excess volatility of prices relative to fundamentals. Hence a key determinant of price volatility is the composition of trading strategies within the population of speculators.

One can take this analysis a step further by observing that the profit accruing to any given trading strategy itself depends on the stability of the market. Fundamental analysis involves considerable time, effort and expense relative to technical analysis. In stable markets, where prices reliably track fundamentals, technical strategies can effectively extract fundamental information from market data at a fraction of the cost of acquiring it directly. Provided that they can process and respond to this information with sufficient speed, such strategies will tend to proliferate when sufficiently rare. Eventually they become too widespread for price stability to be maintained. The result is a transition to an environment with increased volatility in prices relative to the volatility of measures of fundamentals.

It is in this high volatility environment that the value of fundamental analysis is greatest, since departures of prices from intrinsic values arise with greater frequency and on a larger scale. Over the short term, trading on fundamentals can result in losses, especially if positive feedback investment strategies cause an asset price bubble to form. As long as the bubble continues to expand, technical strategies prosper and increase their share of notional market wealth. But asset price inflation in excess of earnings growth cannot be sustained indefinitely, and when the correction comes there is a sharp redistribution of wealth in favor of fundamental strategies. The result is a return to stability in the price dynamics and a corresponding decline in volatility, at which point the process begins anew. The underlying, largely unobservable, changes in market ecology have an observable counterpart in the alternation between high and low volatility regimes.[1]

This perspective on asset price dynamics provides a very natural explanation for several empirically documented features of speculative markets: the excess volatility of prices relative to fundamentals, the clustering of asset price volatility, and the coexistence of short-run momentum with fundamental reversion over longer horizons.[2] All of these phenomena suggest that the volatility of prices does not simply reflect the volatility of the steams of income to which asset holders have claim; some volatility is introduced by the *trading process* itself. This has some major implications for economic growth and welfare. Since the risk-bearing capacity of an economy is not unbounded, increased volatility in major asset classes such as stocks and bonds can result in portfolio shifts towards safer assets such as Treasury bills. This constrains the financing of riskier real projects that have the greatest potential for generating output and employment growth.

Recent changes in technology and regulation have substantially lowered the costs of technical analysis and made feasible virtually instantaneous responses to incoming market data using powerful computers located in close proximity to exchanges. One interpretation of the flash crash is that the proliferation of these strategies took the system past a bifurcation point at which stability of the price dynamics was lost. The algorithmic implementation of such strategies gave the event a peculiarly modern flavor, and resulted in trades at prices that would have been inconceivable in an age of traditional market makers. This, together with the fact that the instability resulted initially in a sharp fall rather than a sudden rise in prices made the mispricing of securities obvious and resulted in a rapid recovery. Had the instability been more modest and in the upward direction, it may well have been subject to more ambiguous interpretation, and taken much longer to correct.

Unlike the market makers of old, high-frequency trading firms do not enjoy substantial monopoly power; they are subject to relatively free entry and intense competition. Despite this, they have managed to secure substantial and sustained profits over the course of their existence.[3] How can one reconcile this level of performance with the competitive structure

[1] This process of endogenous regime switching is described more formally in Sethi (1996); see Brock and Hommes (1997) and Lux (1998) for related models, and Hommes (2006) for a comprehensive survey.

[2] See Shiller (1981) for evidence on excess volatility, Engle (1982) on volatility clustering, DeBondt and Thaler (1985) on fundamental reversion, and Jegadeesh and Titman (1993) on short-run momentum. Hong and Stein (1999) and Abreu and Brunnermeier (2003) develop models that account for the latter two phenomena.

[3] Tradebot, among the largest of the high-frequency trading firms, allegedly managed a four year stretch without a single losing day while having an average holding period of just 11 seconds (Creswell, 2010).

of the market in which these firms operate? One possibility is that their strategies expose them to highly asymmetric payoff distributions involving significant tail risk. They make steady profits with high probability but are exposed to a significant loss if a low probability event such as a major market disruption were to arise. Greater entry increases the likelihood of such a disruption, but profits continue to accrue until the event actually occurs.[4]

The tendency of strategies which carry significant tail risk to proliferate over time even as this risk increases is a phenomenon that applies not simply to technical analysis in speculative markets, but to financial practices more generally. Understanding this process helps explain how some of the most storied names in the financial sector found themselves facing imminent bankruptcy in 2008, bringing the entire system to the brink of collapse.

Tail Risk, Leverage and Maturity Transformation

Among the defining moments of the recent financial crisis was the failure of Lehman Brothers on September 15, 2008, and the rescue of the American International Group (AIG) a day later. AIG was unable to meet collateral calls on credit default swaps written by its financial products subsidiary. These insurance-like contracts promise payments to counterparties in the event of a default by issuers of debt. They can be used to insure against default by individual firms or states, and also against defaults that affect payments to holders of structured products such collateralized debt obligations.[5]

AIG was particularly exposed to the risk of an increase in mortgage default rates sharp enough to affect payments to holders of senior tranches of collateralized debt obligations. Not only had the company written contracts to insure against such default, it also held substantial quantities of the kind of assets it was insuring, which further increased its vulnerability to a sharp rise in mortgage delinquencies (Stulz, 2010). The company had accumulated a position over time that entailed a large directional bet against widespread mortgage default.

[4]The finding by the Kirilenko team that the fastest moving firms were able to exit their positions and shut down operations during the flash crash while significant losses were inflicted on those who were slower to react is consistent with this interpretation.

[5]Collateralized debt obligations are tiered claims to the revenues generated by bundled securities such as bonds and mortgages. Senior tranches are affected by default on the underlying securities only if all subordinate tranches have been completely wiped out; this allows them to carry much higher credit ratings (Coval et al., 2010). Credit default swaps differ from standard insurance contracts in one important respect: they do not require the buyers of protection to have an insurable interest, and can therefore be purchased without any exposure to the underlying credit risk. Hence they allow speculators to bet on default more easily and with greater leverage than they could by short selling bonds.

The payoffs from this bet were highly asymmetric: with high probability it would receive periodic payments from buyers of protection (as well as revenues from its own holdings of mortgage backed securities), and with low probability would face massive losses.

It is tempting to view the accumulation of such a position as reflecting poor risk management or reckless behavior, and in retrospect it was certainly viewed as folly. But from an ecological perspective, the emergence from time to time of large financial intermediaries with major exposure to tail risk of this kind is an inevitable consequence of market competition. At any point in time there is a heterogeneity of beliefs regarding the likelihood of any given rare event. Those who are most convinced that this probability is negligible will offer the best terms for insurance against the event. Their direct competitors will be forced to match these terms or to accept losses in market share. As long as the low probability event does not occur, the highest profits will accordingly accrue to those with the most optimistic beliefs about the assets that they are insuring. As they expand their market share, and others reach for similar levels of performance, prices will increasingly come to reflect these beliefs.

In retrospect, we see that AIG was among the firms most aggressively betting against widespread mortgage default. But had it not been for this firm, some other financial intermediary would have been at the extreme end of the belief distribution. The cost of protection would have been somewhat higher, and the risk exposure more modest, but the high returns from the strategy would have drawn others in, with predictably similar consequences.

While credit derivatives are a relatively recent product of financial engineering, a similar dynamic applies to simpler debt contracts. The term structure of interest rates is typically such that the cost of borrowing rises with the maturity of debt, reflecting differences in the preferred habitat of borrowers and lenders.[6] Lenders prefer shorter maturities while many borrowers need to issue long term debt. This opens up a profit opportunity for financial intermediaries who can engage in maturity transformation by borrowing short and lending long. Holding long-dated assets funded by short-dated liabilities requires the intermediary to roll over its debt periodically, and subjects it to the risk that such financing will not be forthcoming when needed, or will only be available at punitive rates.

A widespread failure by solvent firms to roll over short-term debt is is an unlikely event under normal circumstances, but becomes increasingly likely as the level of maturity transformation in the economy increases. Such increases in maturity transformation arise en-

[6]That is, the yield curve is rising on average, although its steepness and general shape are affected by expectations regarding the future course of interest rates.

dogenously under pressure of market competition. As long as credit markets allow for short term debt to be rolled over by solvent borrowers, the highest profits accrue to those with the greatest maturity mismatch. This drives others to increasing levels of maturity transformation, until a fear of being unable to roll over short term debt triggers a crises of liquidity. Short-term funding dries up just when it is most in demand. If enough firms are forced to unload long-dated assets into a falling market at fire sale prices, large numbers of firms can simultaneously become insolvent.[7]

One the eve of the financial crisis, maturity transformation had reached alarming levels. Overnight loans were secured by collateralized debt obligations in the repo market on an unprecedented scale. In effect, money market funds were being used to finance the purchase of mortgage backed securities. The inability of Lehman to continue to borrow overnight precipitated its collapse and caused Reserve Primary, a money market fund exposed to Lehman commercial paper, to stop redeeming shares at par. This threatened a run on money market funds, the prevention of which was a major motivating factor behind the bailout of AIG and subsequent interventions by the Treasury and the Federal Reserve.[8]

As in the case of protection sellers, banks were driven under pressure of market competition to engage in maturity transformation on an ever increasing scale. As long as the crisis was held at bay, the greatest profits accrued to banks adopting the most aggressive financial practices, and others were compelled to follow. In the memorable words of Citigroup chief executive Chuck Prince, "as long as the music is playing, you've got to get up and dance" (Nakamoto and Wighton, 2007).

What is true of the behavior of banks is true also of the financial practices of households and non-financial firms. At the heart of economic booms and busts, fueled by competitive pressure, lie procyclical movements in leverage. In a recent paper on the subject, John Geanakoplos (2010) notes that in ebullient times "competition drives leverage higher and higher," while during a crisis, "leverage can fall by 50% overnight, and by more over a few days or months." Such large swings in leverage can swamp the effects on output and employment of traditional monetary policy instruments such as interest rates.

[7]This process is vividly illustrated by the case of Northern Rock, the first British bank to fail in 150 years. The bank was not engaged in subprime lending, and had securitized and sold a major portion of its loans. But the loans remaining on its balance sheet were funded by short term debt raised and rolled over in the money market. The bank was not a casualty of high rates of mortgage delinquency or large scale withdrawals by retail depositors, but rather an inability to roll over its debts when markets froze in the face of a spike in the demand for liquidity (Shin, 2009).

[8]See Soros (2009) for a contemporaneous account of these events.

The centrality of leverage and maturity transformation in understanding economic fluctuations was a recurring theme in the work of Hyman Minsky.[9] In Minsky's theoretical framework, firms (and households) are viewed much like banks, as owners of assets that generate cash flows and liabilities that require settlement over a sequence of periods. At any point in time, a firm anticipates a flow of revenues from its productive operations, and faces a flow of contractual debt obligations based on its borrowing history. A crucial distinction is made between "hedge" and "speculative" financing units. Hedge units expect to receive cash flows in each period that are sufficient to cover debt repayments as they come due. As long as these expectations are realized, there is no need for debt to be rolled over. Hence changes in interest rates can affect the value of the firm but cannot threaten it with insolvency. Speculative units, in contrast, anticipate a shortfall of revenues relative to debt obligations in some periods. They expect to have to raise new funds in order to repay debts as they come due, and are therefore vulnerable to increases in interest rates or the unavailability of financing. Furthermore, a sufficiently large increase in their cost of borrowing can lead to a present value reversal and the threat of insolvency.[10]

Like banks engaged in maturity transformation, speculative units use short term debt to finance assets that yield revenues far into the future. The only essential difference is that the assets are real rather than financial. These firms secure a lower cost of capital at the expense of greater exposure to rollover risk. As long as the economy remains dominated by hedge financing units, debts can be rolled over without difficulty. But this leads to an endogenous increase in the incidence of speculative financing (Minsky, 1982, p.65):

> The natural starting place for analyzing the relation between debt and income is to take an economy with a cyclical past that is now doing well. The inherited debt reflects the history of the economy, which includes a period in the not too distant past in which the economy did not do well. Acceptable liability structures are based upon some margin of safety, so that expected cash flows, even in periods when the economy is not doing well, will cover contractual debt payments. As the period over which the economy does well lengthens, two things become evident in board rooms. Existing debts are easily validated and units that were heavily in debt prospered; it paid to lever. After the event it becomes apparent that the

[9]See also Kindleberger (1978), Volcker (1979), and Guttentag and Herring (1984) for similar perspectives on economic fluctuations.

[10]Minsky also introduces a third category, which he calls Ponzi units, defined as firms whose anticipated revenues are insufficient to cover even the interest component of their obligations. Distinguishing between speculative and Ponzi financing units is not essential to the argument here.

> margins of safety built into debt structures were too great. As a result, over a period in which the economy does well, views about acceptable debt structure change. In the deal making that goes on between banks, investment bankers, and businessmen, the acceptable amount of debt to use in financing various types of activity and positions increases. This increase in the weight of debt financing raises the market price for capital assets and increases investment. As this continues, the economy is transformed into a boom economy.

The increase in speculative relative to hedge financing has two effects: it facilitates rapid economic expansion, while also increasing financial fragility. Expansion is facilitated because speculative finance allows firms to grow faster than they could if their borrowing was constrained by expected revenues in each period. Fragility rises because of an increasing need for refinancing as revenues fall short of obligations for a growing number of firms.

The boom is interrupted by a crisis when two conditions are satisfied: a large number of investments are found to be inept, so that cash flows fall below expectations for many firms at once, and financial fragility has reached levels that make it difficult for all those seeking refinancing to find acceptable terms that maintain solvency. An acute need for refinancing arises at a time when banks are reluctant to roll over existing debt. The result is higher interest rates, credit rationing, present-value reversals, strong demand for liquidity, and a collapse in asset prices. If the government sector is large, automatic stabilizers and fiscal policy intervention can prevent a collapse in business profits. Similarly, aggressive monetary expansion can mitigate the extent of credit rationing and the fall in asset prices. But both policies come with a cost: the former in maintaining an inefficient industrial structure and the latter in generating expectations that the risks associated with speculative financing are not severe. The short term benefits of intervention (which are substantial) give rise to long term costs in terms of the frequency and severity of future crises.

Expectations play a central role in this analysis. There is a sense in which expectations, rather than being self-fulfilling as in standard theory, are *self-falsifying*. A crisis occurs when liability structures cause the system to be fragile, but these liability structures are based on optimistic beliefs about the prospects for refinancing. Similarly, pessimism about rollover risk results in robust liability structures that facilitate refinancing on favorable terms. This is the basis for Minsky's paradoxical claim that stability is itself destabilizing (1975, p.126). Firms persist in adopting liability structures which give rise to outcomes which violate the assumptions on the basis of which the liability structures were chosen. This occurs both in the case of excessive caution following a period of instability, and excessive boldness following a long expansion.

Minsky's assumptions regarding expectation formation have attracted some criticism. Flemming (1982), for instance, claims that "the argument depends on agents failing to distinguish a run of good luck from a favorable structural shift in their environment."[11] From the perspective of market ecology, such criticisms miss the point. The expectations that come to be reflected in liability structures are those held by individuals whose actions have recently been rewarded by success. Over the course of a stable expansion, success accrues in greatest measure to those with the most aggressive financial practices. Those who anticipate a crisis and pad their margins of safety are punished, not rewarded, as long as the stable expansion continues.

More generally, the distribution of forecasting rules and procedures in the population of firms is endogenously determined under pressure of market competition. In the terminology of Marcet and Sargent (1989), the economy is a self-referential system: the objective laws of motion governing the evolution of economic magnitudes depend on the perceived laws of motion on the basis of which individuals act. It is commonplace to assume in contemporary macroeconomic modeling that the perceived laws coincide with the actual laws to which they give rise. As noted above, this so-called "rational expectations" hypothesis is in fact a hypothesis of consistent or equilibrium expectations. It is certainly the case that the agents in Minsky's model do not have equilibrium expectations in this sense. But it is also conceivable that such equilibrium expectations are locally unstable under plausible disequilibrium dynamics based on experience-based learning or evolutionary selection among heterogeneous beliefs. This is a question to which future research effort could fruitfully be directed.[12]

Expectations

Foley and Sidrauski left open the question of how expectations are determined, just as Keynes had done a generation earlier. The rational expectations hypothesis sidestepped the problem altogether by simply assuming that expectations were determined by equilibrium conditions once other primitives of a model (preferences, endowments, technologies and policy rules for instance) had been specified. As noted by Woodford (2011), however, expectation formation

[11] Similar concerns have been expressed about the work of Kindleberger; see, for instance, Bernanke (1983).
[12] For a more formal statement of these points and some preliminary analysis, see Sethi (1992, 1995).

and revision as an "object of analysis in its own right" is now an area of active research interest in macroeconomics. One approach is that of learning.

A substantial literature on learning in macroeconomics has emerged over the past couple of decades, but much of it considers a representative agent attempting to learn the value of one or more parameters.[13] The prototypical model in this literature is based on the assumption that the dynamics of state variables have a recursive structure that is consistent with a rational expectations solution, but with a parameter that is unknown to the representative agent. Learning changes beliefs about the parameter, but also changes the parameter itself since it is sensitive to agent beliefs. Convergence occurs when beliefs about the parameter match the value of the parameter that these beliefs induce, at which point learning ceases.

Even within this simple framework, learning does not always converge to rational expectations. Howitt (1992) provided a very clear example of this in his analysis of a nominal interest rate peg. Starting from a steady state in which expectations are fulfilled, a permanent decline in the nominal interest rate results in an immediate transition to lower levels of inflation under rational expectations. But under a broad class of learning dynamics, such a decline instead results in accelerating inflation, just as Wicksell's (1898) analysis of the cumulative process would predict. Howitt concludes as follows:

> Perhaps the most important lesson of the analysis is that the assumption of rational expectations can be misleading, even when used to analyze the consequences of a fixed monetary regime. If the regime is not conducive to expectational stability, then the consequences can be quite different from those predicted under rational expectations... in general, any rational expectations analysis of monetary policy should be supplemented with a stability analysis... to determine whether or not the rational expectations equilibrium could ever be observed.

Such models are suggestive, but do not really tackle the question of the manner in which the decentralized and uncoordinated intertemporal plans of large numbers of households and firms are brought into consistency with each other. To do so requires a multi-agent framework with heterogeneous beliefs and forecasting rules in competition with each other. Realized values of economic magnitudes depend on the distribution of such rules in the population. The performance of the various rules depends on the frequencies with which they are present

[13] See Marcet and Sargent (1989) and Evans and Ramey (1992) for early contributions, and Evans and Honkapohja (1999) for a comprehensive survey.

in the population, and these frequencies evolve under pressure of differential payoffs. The resulting dynamics of beliefs and realizations can then be studied.

This is the methodology of agent-based computational economics.[14] Recent work by Blake LeBaron on belief and price dynamics in a speculative asset market illustrates the value of this approach. In LeBaron's model, a broad range of forecasting rules, including those associated with fundamental and technical trading strategies, compete with each other (and with a passive, belief-independent strategy). The resulting price dynamics display two characteristics commonly found in high frequency data on asset returns: short-run momentum and long-run mean reversion. Furthermore, instead of convergence to anything that might resemble homogenous and self-fulfilling beliefs, a broad range of forecasting rules continue to co-exist indefinitely.

The agent-based approach to learning and evolutionary selection has the potential to provide microeconomic underpinnings for a number of models that have languished since the rational expectations revolution. In particular, it could allow us to revisit from a fresh perspective theories of economic fluctuations that appear to be inconsistent with intertemporal optimization and self-fulfilling expectations, but which might well be consistent with competition among heterogeneous forecasting rules. Some of the richest and most interesting endogenous business cycle theories were developed without assuming that agent expectations are self-fulfilling, and indeed without much concern for microeconomic foundations at all. Local instability in such models is often caused by accelerator effects in the goods market, and a variety of different nonlinear effects that keep trajectories bounded have been considered.[15]

The perspective of market ecology can allow for a renewed exploration and reinterpretation of this work, which is based on sound economic logic but lacks clarity at the level of individual decision making. Agent-based computational methods could prove especially useful in this regard.[16] But there is an important obstacle to the widespread adoption of this methodology in economics, and it concerns the issue of robustness. Theoretical models

[14] Numerous applications of this approach to problems in economics are surveyed in Tesfatsion and Judd (2006); see Farmer and Foley (2009) for a recent article advocating its broader use. Among the earliest and most effective uses of this method in the social sciences may be found in Thomas Schelling's (1971) analysis of segregation.

[15] Samuelson (1939), Kaldor (1940), Hicks (1950), Goodwin (1951), Leijonhufvud (1973), Tobin (1975), and Foley (1987) are important examples. This research tradition, based on aggregative models of disequilibrium dynamics, continues in the recent work of Chiarella et al (2005).

[16] Gatti et al. (2008) take an important step in this direction.

that yield analytical solutions can explicitly characterize the set of function specifications and parameter values for which a given result holds. Numerical models, in contrast, can only explore some finite subset of this space. In the former case the question of model robustness is addressed by asking whether the theoretical claims remain valid under small changes in specification. This question is typically impossible to answer definitively on the basis of simulations. As a result, there is no established criterion for judging the quality of an agent-based model. Once a professional consensus on this issue has emerged, the floodgates will open to research in this area. Until then, we shall continue to ask the question first posed by Thorstein Veblen in 1898: "Why is Economics not an Evolutionary Science?"

References

[1] Beja A. and M.B. Goldman (1980). "On the dynamic behavior of prices in disequilibrium." Journal of Finance 35: 235–248.

[2] Benhabib, J. and R.H. Day (1982). "A characterization of erratic dynamics in the overlapping generations model." Journal of Economic Dynamics and Control 4: 37–55.

[3] W.A. Brock and C.H. Hommes (1997). "A rational route to randomness." Econometrica 65: 1059–1095.

[4] C. Chiarella, "The dynamics of speculative behaviour." Annals of Operations Research 37: 101–123.

[5] Chiarella, Carl, Peter Flaschel, and Reiner Franke (2005). Foundations for a disequilibrium theory of the business cycle: qualitative analysis and quantitative assessment Cambridge University Press.

[6] Creswell, Julie (2010). "Speedy New Traders Make Waves Far From Wall St." New York Times, May 16.

[7] Engle, Robert F. (1982). "Autoregressive conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." Econometrica 50: 987–1008.

[8] Evans, George and Garey Ramey, (1992) "Expectation Calculation and Macroeconomic Dynamics." American Economic Review 82: 207-224

[9] Evans, George and Seppo Honkapohja (1999). "Learning dynamics." In: John B. Taylor and Michael Woodford, Editor(s), Handbook of Macroeconomics, Elsevier, 1999, Volume 1, Part 1, pp. 449–542.

[10] Farmer, J. Doyne and Duncan Foley (2009). "The economy needs agent-based modelling." Nature 460: 685–686.

[11] Foley, Duncan K. (1999) "The Ins and Outs of Late Twentieth Century Economics." In Arnold Heertje (ed.), The makers of modern economics, Volume 4, E. Elgar.

[12] Foley, Duncan K. (1987). "Liquidity-profit rate cycles in a capitalist economy." Journal of Economic Behavior & Organization 8: 363–376.

[13] Foley, Duncan K. and Miguel Sidrauski (1971). Monetary and fiscal policy in a growing economy, Macmillan.

[14] Flemming, J.S. (1982): ”Comment.” In Financial Crises: Theory, History, and Policy, edited by C. P. Kindleberger and J.-P. Laffargue. Cambridge University Press.

[15] Gunter Gabisch, Hans-Walter Lorenz (1987). Business cycle theory: a survey of methods and concepts, Springer-Verlag.

[16] Gatti, D., Gaffeo, E., Gallegati, M., Giulioni, G., Palestrini, A. (2008). Emergent Macroeconomics: An Agent-Based Approach to Business Fluctuations, Springer.

[17] Geanakoplos, John (2010). “The Leverage Cycle.” NBER Macroeconomics Annual.

[18] Goodwin, Richard M. (1951). “The Nonlinear Accelerator and the Persistence of Business Cycles.” Econometrica 19: 1–17.

[19] Goodwin, Richard, and Lionello F. Punzo (1987). The dynamics of a capitalist economy: a multi-sectoral approach, Westview Press.

[20] Grandmont, J.-M. (1985). On endogenous competitive business cycles. Econometrica 53: 995–1046.

[21] Grant, Jeremy and Michael Mackenzie, 2010, “Markets: Ghosts in the machine.” Financial Times, February 17.

[22] Hicks, John (1950). A contribution to the theory of the trade cycle, Clarendon Press.

[23] Hommes, Cars H. (2006). “Heterogeneous Agent Models in Economics and Finance.” In Handbook of Computational Economics Volume 2, pp. 1109-1186.

[24] Howitt, Peter (1992). “Interest Rate Control and Nonconvergence to Rational Expectations.” The Journal of Political Economy 100: 776–800.

[25] Kaldor, Nicholas (1940). “A Model of the Trade Cycle.” The Economic Journal 50: 78–92.

[26] Keynes, John Maynard (1936). The General Theory of Employment, Interest, and Money.

[27] Kindleberger, C. P. (1978): Manias, Panics, and Crashes: A History of Financial Crises. New York: Basic Books.

[28] Kirilenko, Andrei A., Kyle, Albert S., Samadi, Mehrdad and Tuzun, Tugkan (2011). "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market." Available at SSRN: http://ssrn.com/abstract=1686004

[29] LeBaron, Blake (2010). Heterogeneous Gain Learning and the Dynamics of Asset Prices. Unpublished Manuscript, Brandeis University.

[30] Leijonhufvud, Axel (1973). "Effective Demand Failures." The Swedish Journal of Economics 75: 27–48.

[31] Lux, Thomas (1998). "The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions." Journal of Economic Behavior & Organization 33: 143–165.

[32] Mandelbrot, Benoit B. 1963. "The Variation of Certain Speculative Prices." The Journal of Business 36: 394-419.

[33] Marcet, Albert and Thomas J. Sargent (1989). "Convergence of least squares learning mechanisms in self-referential linear stochastic models." Journal of Economic Theory 48: 337–368.

[34] Minsky, Hyman (1975). John Maynard Keynes. New York: Columbia University Press.

[35] Minsky, Hyman (1982). Can "It" Happen Again: Essays on Instability and Finance. Armonk, NY: M.E. Sharpe.

[36] Minsky, Hyman (1986). Stabilizing an Unstable Economy. New Haven, CT: Yale University Press.

[37] Nakamoto, Michiyo and David Wighton (2007). "Citigroup chief stays bullish on buyouts." Financial Times, July 9.

[38] Samuelson, Paul A. (1939). "A Synthesis of the Principle of Acceleration and the Multiplier." The Journal of Political Economy 47: 786–797.

[39] Schelling, Thomas C. (1971). "Dynamic models of segregation." The Journal of Mathematical Sociology 1: 143-186.

[40] Sethi, Rajiv (1992). "Dynamics of learning and the financial instability hypothesis." Journal of Economics 56: 39–70.

[41] Sethi, Rajiv (1995) "The Evolutionary Dynamics of Financial Practices." Metroeconomica 46: 246–277.

[42] Sethi, Rajiv (1996). "Endogenous regime switching in speculative markets." Structural Change and Economic Dynamics 7: 99–118.

[43] Shiller, Robert J. (1981). "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" The American Economic Review 71: 421–436.

[44] Shin, Hyun Song (2009). "Reflections on Northern Rock: The Bank Run That Heralded the Global Financial Crisis." Journal of Economic Perspectives 23: 101–19.

[45] Soros (2009). "The game changer." Financial Times , January 28.

[46] Stultz, Rene M. (2010). "Credit Default Swaps and the Credit Crisis." Journal of Economic Perspectives 24: 73–92.

[47] Tesfatsion, L. and K.L. Judd, Editor(s), Handbook of Computational Economics, Elsevier, 2006, Volume 2, Pages 831-880.

[48] Tobin, James (1975). "Keynesian Models of Recession and Depression." The American Economic Review Papers and Proceedings 65: 195-202.

[49] Veblen, Thorstein (1898). "Why is Economics not an Evolutionary Science?" The Quarterly Journal of Economics 12: 373-397.

[50] Volcker, Paul (1979). The Rediscovery of the Business Cycle. Free Press.

[51] Woodford (2011). "A Response to John Kay." INET Blog, October 5.

[52] Wicksell, Knut (1898). Interest and Prices, translated by R.F. Kahn (1936), London: Macmillan.