Models of Markets: Finance Theory and the Historical Sociology of Arbitrage:

Donald MacKenzie

January 2004
Author’s address:
School of Social and Political Studies
University of Edinburgh
Adam Ferguson Building
Edinburgh EH8 9LL
Scotland
D.MacKenzie@ed.ac.uk
Abstract

The mathematical models of modern financial economics are not simply “academic”: they are used by market practitioners, and influence both the overall shape of markets and particular chains of events. Critical to many such models is the assumption that if empirical prices differ from the values they “should” have (according to the model), then the resultant price discrepancies will be exploited by arbitrageurs and thus eliminated or reduced. This paper surveys the significance of arbitrage for finance theory’s models (particularly in relation to the pricing of derivatives), examines the practice of arbitrage (via a case-study of the arbitrageurs, Long-Term Capital Management) and argues that “arbitrage” is a topic for the historian and sociologist as well as for the economist.
Models of Markets: Finance Theory and the Historical Sociology of Arbitrage

Donald MacKenzie*

In few areas of economics and of the economy are mathematical models of greater significance than in finance.¹ Over the last fifty years, the academic study of finance has been transformed from a largely descriptive, non-mathematical enterprise to a highly analytical one in which sophisticated mathematics is deployed and for which Nobel Prizes in economics have been

* Donald MacKenzie, School of Social and Political Studies, University of Edinburgh

¹ I am extremely grateful to the Revue’s referees for extremely helpful comments on the original version of this paper, which was presented to the Colloque International, Centre Alexandre Koyré, “Modèles et Modélisations, 1950-2000: Nouvelles Pratiques, Nouveaux Enjeux,” and to London School of Economics, Centre for the Analysis of Risk and Regulation, Workshop on “Organisational Encounters with Risk.” Parts of this paper will also appear in a book with the latter title, edited by Bridget Hutter and Michael Power (Cambridge University Press). The research reported here was supported originally by DIIRC, the Interdisciplinary Research Collaboration on the Dependability of Computer-Based Systems (UK Engineering and Physical Sciences Research Council grant GR/N13999). It is being continued with the support of a professorial fellowship awarded by the UK Economic and Social Research Council (RES-051-27-0062).
awarded (see, e.g. Bernstein 1992, Bouleau 1998). The resultant models have moved from academia into the markets themselves, where they are widely used by investment professionals. In particular, the models have been critical to a key transformation: the emergence of a huge global market in financial derivatives. (A “derivative” is an asset the value of which depends upon the price of another “underlying” asset or on the level of an index or interest rate. Options – to be discussed below – are examples of derivatives.) As recently as 1970, the markets in financial derivatives were tiny, and many derivatives were illegal. By December 2002, derivatives contracts totaling US $166 trillion were outstanding worldwide, a sum equivalent to around $27000 for every human being on earth.² It would be difficult in the extreme to trade these trillions of dollars of derivatives without a guide to how to price them and how to hedge the risk they involve. Modern finance theory has provided this guide.

The role of mathematical models in this twin transformation – of the academic study of finance and of the financial markets – has many aspects. This article focuses on only one of those aspects: arbitrage. Arbitrage is trading

²Data from the Bank for International Settlements http://www.bis.org This figure arguably overstates the economic significance of derivatives, for example by valuing swaps (see note 17 below) by total notional principal sums, when the principal does not actually change hands. Nevertheless, derivatives trading remains a major activity even if $166 trillion is deflated by a factor of ten or even 100.
that exploits price discrepancies. It is a key mechanism – arguably the key mechanism – in mathematical models of financial markets. In these models, it is above all arbitrage that is invoked to guarantee that the pattern of prices will correspond to the model. In models of the pricing of derivatives, for example, the most common form of reasoning is that the only price pattern that can be stable is one in which there are no opportunities for riskless arbitrage profits. When combined with other assumptions – for example about the stochastic dynamics of the price of the underlying asset – this reasoning frequently gives rise to an equation or equations that can be solved either analytically or numerically to yield the price of the derivative.

Arbitrage is also, of course, an important form of trading in actual markets. To be sure, the economist’s theoretical notion of arbitrage (trading that demands no net capital and yields a profit without incurring risk) does not correspond exactly to what market practitioners mean by the term, since they frequently count as “arbitrage” forms of trading that do demand capital and involve risk. The relation between “arbitrage” as a mechanism in mathematical models and “arbitrage” as market practice is indeed the central theme of this article, which has five parts. After this introduction comes a section discussing in more detail the role of arbitrage in models of financial markets. The third section discusses the relationship between these models and “reality,” alluding

---

3 A brief history of financial economics that (correctly) emphasizes arbitrage is Harrison (1997).
for example to Michel Callon’s claim (Callon 1998) that economics is performative: that it brings into being the phenomena it describes. The fourth section presents a case-study of the arbitrageurs of the hedge fund Long-Term Capital Management (LTCM), who received widespread publicity when, in September 1998, the fund nearly became bankrupt and was recapitalized by a consortium of the world’s leading banks brought together by the Federal Reserve Bank of New York. The case of LTCM has many aspects: it can, for example, be used as a lesson in the importance of “fat tails” (the tendency for extreme events to occur in financial markets far more frequently than implied by standard mathematical models of the stochastic dynamics of prices) or of the tendency in a crisis for the correlations of the prices of apparently unrelated financial assets to rise dramatically and dangerously. However, the most central lesson of the LTCM case, from the viewpoint of this article, lies in what it reveals about the contingencies of “arbitrage” as market practice. The final section of the article argues that those contingencies are historical and sociological as well as economic. At least in relation to arbitrage, therefore, an understanding of the relations of models to “reality” in the financial markets require the perspectives of the historian and sociologist as well as those of the economist.

**Arbitrage and Models of Financial Markets**

The invocation of arbitrage as a mechanism in models of financial markets can
be traced most centrally to the Nobel-prize-winning work of Franco Modigliani and Merton Miller, who demonstrated that in a “perfect market” (Modigliani and Miller 1958, p. 268) the total value of a firm is not affected by its “capital structure,” that is, by its degree of leverage, the extent to which it chooses to finance its activities by the issuance of debt such as bonds rather than stock. What was of significance was not just Modigliani and Miller’s proposition, but the way they proved it: a way that has become known as “arbitrage proof.” They showed that if two firms with different capital structures but identical expected future income streams were valued differently by the market, “arbitrage will take place and restore the stated equalities.” In other words, “an investor could buy and sell stocks and bonds in such a way as to exchange one income stream for another stream, identical in all relevant respects but selling at a lower price. ... As investors exploit these arbitrage opportunities, the value of the overpriced shares will fall and that of the underpriced shares will rise, thereby tending to eliminate the discrepancy between the market values of the firms” (Modigliani and Miller 1958, p. 269).

The key papers initiating the use of “arbitrage proof” to determine the price of derivatives were by Fischer Black, Myron Scholes, and Robert C. Merton (Black and Scholes 1973; Merton 1973), work that in 1997 won Nobel Prizes for Scholes and Merton (Black died prematurely in 1995). The problem solved by Black, Scholes, and Merton was the pricing of options: contracts that
confer the right, but not the obligation, to buy (“call”) or sell (“put”) a given asset at a given price, at (or up to) a given expiry date. Again assuming a perfect market (no transaction costs and, for example, the capacity both to borrow and to lend at the riskless rate\(^4\) of interest), they showed that an option on an asset such as stock could be replicated completely by a continuously-adjusted portfolio of the asset and cash, so long as the returns on the asset followed the by-then-standard model of a log-normal\(^5\) random walk in continuous time.\(^6\) If the price of the asset diverges from the cost of the replicating portfolio, arbitrageurs will buy the cheaper and short sell\(^7\) the dearer

---

\(^4\) The “riskless rate” is the rate of interest paid by a borrower whom lenders are certain will not default. The yield of bonds issued in their own currency by major governments is typically taken as an indicator of the riskless rate.

\(^5\) If stock prices themselves were normally distributed, there would be a non-zero probability of negative prices, and limited liability means that stock prices cannot be negative. Log-normality of price changes was a more attractive assumption because it avoided this problem (a variable is log-normal if its logarithm follows a normal distribution).

\(^6\) I am here oversimplifying a complex historical development: see MacKenzie (2003b).

\(^7\) To “short sell” or “short” an asset is to sell an asset one does not own, for example by borrowing it, selling it, and later repurchasing and returning it.
of the two, and they will continue to do so until equality is restored. More

generally, Black’s, Scholes’s, and Merton’s analyses suggested a methodology

for the rational pricing and hedging of derivative products of all kinds: identify

the replicating portfolio of more basic assets (if it exists), and use its cost to

work out the price of the derivative, and (if desired) to hedge its risks. This

methodology is, as noted in the introduction, key not just to the academic study

of derivatives but also to the practice of derivatives markets.

Commentary on LTCM’s crisis has often drawn a connection between

the events of August and September 1998 and the assumption of log-normality

in Black-Scholes-Merton option pricing: some of the price movements of those

months were indeed wildly improbable on the hypothesis of log-normality (see

MacKenzie 2003a). To focus upon log-normality, however, is to focus on a less-

than-central aspect of Black, Scholes, and Merton’s contribution to finance

theory: that stock price changes were not in practice log-normal was known

even in 1973, when their work was published. As Bouleau (1998, p. 63) puts it,

the “epistemological rupture” is the idea of the replicating portfolio and

consequent possibility of pricing by arbitrage. Merton himself, and other

finance theorists such as Steve Ross, John Cox, Mark Rubinstein, and William
Sharpe, soon showed how to extend the basic framework of Black-Scholes-Merton derivative pricing to worlds in which the dynamics of asset pricing was not log-normal.

The work of Black and Scholes on option pricing was first circulated in October 1970. By 1979, J. Michael Harrison and David M. Kreps had established the form of derivative pricing theory that is most attractive to mathematicians (Harrison and Kreps 1979; Harrison and Pliska 1981). Crucial was the link they drew to the theory of martingales. (A martingale is a stochastic process for which the expected future value of a variable, conditional upon its current value, is its current value. Loosely – there are deep mathematical complications here – a martingale is a “fair game”: in a game of chance which is a martingale, a player’s expectation of gain or loss is zero.) Others had previously realised that financial markets could be modeled as martingales, but it was Harrison, Kreps (and Stanley R. Pliska) who brought to bear the full power of modern martingale theory. Martingale theory freed option pricing from dependence upon any particular stochastic process: it could

---

8 This version is in box 28 of the Fischer Black papers at MIT (Institute Archives, MC505).
encompass the log-normal random walk posited by Black, Scholes, and Merton; the Poisson, “jump,” process investigated by Cox, Ross, and Merton; and the finite-time models of Sharpe, Cox, Ross, and Rubinstein.

Harrison and Kreps showed that in a “frictionless” market with no opportunities for arbitrage, there existed an “equivalent martingale measure,” a way of assigning probabilities to the path followed by the price of an asset such that the arbitrage-imposed price of a derivative contract on that asset was simply the conditional expectation of its payoff discounted back to the present. If the market was complete – in other words, if every contingent claim\(^9\) could be replicated – then the equivalent martingale measure was unique. Harrison and Kreps’s conclusions gave general form to perhaps the most surprising of the findings of the work of Black, Scholes, and Merton: that pricing by arbitrage proof meant that all sorts of complications, notably the degree of risk aversion of investors, could be ignored, and derivatives could be priced as if all investors

\(^9\) A contingent claim (such as an option) is a contract the value of which depends on some future state of the world (for example, the value of an option at its expiry depends on the price of the underlying asset).
were risk-neutral. (To get a sense of what “risk-neutrality” means, imagine being offered a fair bet with a 50% chance of winning $1,000 and a 50% chance of losing $1,000, and thus an expected value of zero. If you would require to be paid to take on such a bet you are “risk averse”; if you would pay to take it on you are “risk seeking”; if you would take it on without inducement, but without being prepared to pay to do so, you are “risk neutral”.)

These 1970s’ developments in derivative pricing theory were greatly elaborated in the 1980s and 1990s, as any textbook of mathematical finance (e.g. Hull 2000) reveals, and, as noted above, were drawn on heavily in the rapidly growing derivatives markets. The theory of derivatives developed by Black, Scholes, and Merton, and added to by Cox, Ross, Rubinstein, Harrison, Kreps, and others, forms an essential part of this huge high-modern industry, guiding participants both in the pricing of derivative products and in hedging the risks involved. The theory and its accompanying mathematical models are built deep into the economic structure of high modernity.

“Arbitrage proof” thus plays a central role in the theory of derivatives.
Arbitrage is also highly significant in the justification of the overall notion of “market efficiency,” which has shaped not just financial economics but also, via the plausibility it lends to notions of rational expectations, economics as a whole, and has helped derivatives markets to grow by providing legitimacy. Options, for example, are not new products: they have been traded since the seventeenth century. They had often been the object of suspicion, however, because they looked dangerously like wagers on price movements. The argument that derivatives could contribute to market efficiency, and were not simply vehicles for gambling, was key to the gradual removal in the 1970s and 1980s of legal barriers to derivatives trading (MacKenzie and Millo 2003).

A financial market is “efficient” if prices in it reflect all available information (Fama 1970). The idea of market efficiency is the key overall foundation of orthodox modern financial economics, as well as contributing to the perceived legitimacy of actual markets. But what might make markets efficient? For some of the central figures in modern financial economics, to assume that all investors are perfectly rational and perfectly well-informed has been just too heroic. It is, for example, difficult on that assumption to explain the high volumes of trading in the financial markets. If markets are efficient, prices already incorporate all publicly available information, and so if all traders are perfectly rational and perfectly well-informed, why should they continue to trade once they have diversified their portfolios satisfactorily? “Noise trading,” said Fischer Black (1986, p. 531) “provides the essential
missing ingredient. Noise trading is trading on noise as if it were information. People who trade on noise are willing to trade even though from an objective point of view they would be better off not trading. Perhaps they think the noise they are trading on is information. Or perhaps they just like to trade.”

If the empirical presence of noise trading and other departures from rationality is hard to deny, and if its denial leads to incorrect predictions (markets with far less trading than in reality), does this then mean that the thesis of market efficiency must be rejected, and some version of “behavioural finance” adopted? Not so, argues Steve Ross (2001, p. 4):

I, for one, never thought that people – myself included – were all that rational in their behavior. To the contrary, I am always amazed at what people do. But, that was never the point of financial theory.

The absence of arbitrage requires that there be enough well financed and smart investors to close arbitrage opportunities when they appear. ... Neoclassical finance is a theory of sharks and not a theory of rational homo economicus, and that is the principal distinction between finance

10 In “behavioural finance,” market participants are assumed to be less than entirely rational, for example to be subject to various systematic biases, normally psychological in their nature.
and traditional economics. In most economic models aggregate demand depends on average demand and for that reason, traditional economic theories require the average individual to be rational. In liquid securities markets, though, profit opportunities bring about infinite discrepancies between demand and supply. Well financed arbitrageurs spot these opportunities, pile on, and by their actions they close aberrant price differentials. ... Rational finance has stripped the assumptions [about the behaviour of investors] down to only those required to support efficient markets and the absence of arbitrage, and has worked very hard to rid the field of its sensitivity to the psychological vagaries of investors.

Performing Theory

Orthodox modern finance economics, including the theory of derivatives, is elegant and powerful. What is the relationship of that theory and its accompanying models to “reality”? They are, of course, an abstraction from it, and known to be such by all involved. Neither finance theorists nor sophisticated practical users of financial models believe in the literal truth of the
models’ assumptions.\textsuperscript{11} Does that lack of verisimilitude mean that, as much of the commentary on LTCM (such as Lowenstein 2000) suggests, the theory is a hopelessly flawed endeavour? Two points suggest not. The first was spelled out by Milton Friedman in his famous essay “The Methodology of Positive Economics” (Friedman 1953). The test of an economic theory, Friedman argued, was not the accuracy of its assumptions but the accuracy of its predictions. That viewpoint has become fundamental not just to modern neoclassical economics but also to finance theory: indeed, one of the distinguishing features of the modern theory of finance is that it abandoned the earlier attitude that the job of the scholar in finance was accurately to describe what people in the finance industry actually did. When the Black-Scholes-Merton option pricing model, for example, was first propounded in the early 1970s, its assumptions were wildly unrealistic. Not only was it already known by then that empirical stock price distributions had “fat tails” (in other words, that the probabilities of extreme events were considerably greater than implied by the log-normal model), but transaction costs were high (not zero as assumed in the model), there were significant restrictions on short-selling stocks, etc. By

\textsuperscript{11} See the interviews drawn on in MacKenzie (2003a, 2003b) and MacKenzie and Millo (2003).
the late 1970s and early 1980s, however, the differences between empirical option prices and best-fit Black-Scholes theoretical prices was remarkably good, with residual discrepancies typically less than 2% (Rubinstein 1985). “When judged by its ability to explain the empirical data,” commented Steve Ross (1987, p. 332), Black-Scholes-Merton option pricing theory and its variants formed “the most successful theory not only in finance, but in all of economics.” (Interestingly, the fit between empirical data and the Black-Scholes-Merton model deteriorated after 1987, but that is a matter, which, for reasons of space must be set aside here: see Rubinstein 1994 and MacKenzie and Millo 2003.)

The second point is that the empirical accuracy of finance theory’s typical assumptions has increased considerably since the 1970s. This is perhaps most apparent in regard to the speed of transactions (the Black-Scholes-Merton option pricing model assumes the possibility of instantaneous adjustment of the replicating portfolio) and transaction costs. Because of technological change and institutional reform (in particular, the abolition of fixed commissions on the New York Stock Exchange and other leading exchanges), for major players in
the main stock markets transaction costs are now close to zero,\textsuperscript{12} and significant adjustments to portfolios can now be made, if not instantaneously, at least in seconds. Finance theory models are still idealizations of market realities, but less radical idealizations than they were in the early 1970s.

Has the crucial assumption that price discrepancies will be closed by arbitrage, like the other assumptions of finance theory, also become more true with the passage of time? The prevalence of arbitrage opportunities is a more difficult point to investigate than, for example, the decline in transaction costs. There is a strong incentive to exploit such opportunities, rather than to reveal them in the academic literature, so they may be underreported. Conversely, however, what may appear to be an arbitrage opportunity may actually disappear as soon as one seeks to exploit it. The typical mechanism by which this happens is what market participants call “slippage”: the movement of prices against one as soon as one starts to trade in significant quantities. Because of slippage and other practicalities, one cannot simply investigate statistically: in a sense, to determine the presence of arbitrage opportunities one must become an arbitrageur. As Ross puts it: “To find the [arbitrage] opportunities one must put oneself in the shoes of the arbitrageurs which is

\textsuperscript{12} Amongst the reasons is that brokers will offer to transact large trades effectively free of commission because of the informational advantages such transactions offer them. Note, however, that slippage (see text) is still a significant issue, and it can also be seen as a transaction cost.
difficult and expensive” (2001, p. 4).

The testimony of actual arbitrageurs (in interview)\textsuperscript{13} is not entirely unequivocal, but all tend to agree that arbitrage opportunities that are relatively easy to identify tend to diminish. The action of arbitrageurs may not close them completely – there may be a kind of “predator-prey” dynamics in which as arbitrage opportunities diminish so too does the commitment of arbitrage capital to exploiting them – and the expansion of the global financial system into new geographical territories and new products creates new opportunities to replace diminished ones. But, overall, it seems reasonable to conclude that in the core financial markets of the Euro-American world the assumption that price discrepancies will be closed by arbitrage has a tendency to become more realistic.

Note that some of the increasing realism of the assumptions of finance theory is due to the very development and acceptance of the theory. As Callon (1998) has pointed out, economic theory has a performative dimension. It does not simply describe an already-existing external world, but can help that world come into being. (The classic study of neoclassical economics creating a market in its own image is Garcia 1986.) Take, for example, option pricing theory. The close empirical fit between the pattern of observed option prices and the Black-

\textsuperscript{13} For these interviews, see note 11.
Scholes-Merton model resulted at least in part from the use of option pricing theory to detect and exploit arbitrage opportunities. The growing prestige of Black-Scholes-Merton theory also directly affected the validity of one of its key assumptions, as Yuval Millo (forthcoming) has discovered. When the theory was initially formulated, its assumption that stocks could be bought entirely on credit was empirically false: in the United States, stock purchases on credit were strictly limited by the Federal Reserve's famous Regulation T. The Black-Scholes-Merton model, however, was used to delineate "bona fide hedges" which were exempted from this restriction and therefore could be implemented using stocks purchased entirely on credit!

If finance theory already had its performative aspects by 1980, by the 1990s it was, as noted above, built into the very fabric of the high-modern financial world. Thus Walter (1996, p. 904) describes the role of Itô's lemma, the

14 The classic arbitrage was "spreading": using option pricing theory to identify relatively "over-priced" and relatively "under-priced" options on the same underlying asset, selling over-priced options and hedging their risk by buying under-priced ones. The effect of the strategy was to push prices in the direction of satisfying the key econometric test (Rubinstein 1985) of the validity of the Black-Scholes-Merton model: that the "implied volatilities" of all the options with the same expiry on the same underlying asset should be identical. This is discussed in more detail in MacKenzie and Millo (2003), which also examines why option prices since 1987 differ from prices prior to 1987 in this respect.
key bridging result between “ordinary” calculus and the stochastic calculus of Merton’s version of option pricing theory and its many more recent developments. Without the lemma, “no trading room could now manage its options market positions.”\textsuperscript{15} The key players in the markets of high-modern finance perform not just general notions of market efficiency but highly sophisticated mathematical formulations. Martingale theory, for example, is no longer simply “pure mathematics,” but is performed in flesh, blood, and silicon in the markets. Finance theory is a world-making, and not just a world-describing, endeavour.

\textit{Long-Term Capital Management}\textsuperscript{16}

LCTM was an investment partnership set up in 1993 by John Meriwether, previously head of Salomon Brothers’ bond arbitrage desk and a senior

\textsuperscript{15} My translation. In fact, there may be rather greater dependence on discrete-time models such as Cox-Ross-Rubinstein than on continuous-time models to which Itô’s lemma applies (see MacKenzie and Millo 2003), but Walter’s generic point is undoubtedly correct.

\textsuperscript{16} This section draws on an earlier treatment of LTCM (MacKenzie 2000), but that treatment is in some respects in error. MacKenzie (2003a) discusses the LTCM case fully – it contains, for example, quantitative tests of the explanation of 1998 that is sketched here – and it should be consulted for details, discussion of sources, etc. In the interests of brevity, I describe here only the outline of the episode.
manager in the bank. Meriwether recruited to LTCM from Salomon and elsewhere an impressive team of experienced traders and specialists in mathematical finance: amongst its partners were Merton and Scholes, whose work on option pricing theory is described above. Much of LTCM’s trading was with leading banks, and it largely avoided risky “emerging markets,” preferring well-established ones such as those in government bonds of the leading industrial nations (though active in the U.S. and Japan, as Salomon had been, LTCM was more heavily involved than Salomon had been in European bond markets), in swaps,\(^\text{17}\) in options, in mortgage derivatives, and in certain very restricted categories of stock.\(^\text{18}\) Following the tradition established by Meriwether at Salomon, the fund eschewed speculation based upon intuitive hunches. It sought pricing discrepancies around which to base arbitrage strategies, and generally constructed its positions so as to be insulated from overall stock market movements and interest rate changes.

LTCM’s market positions were varied, but a common theme underlay many of them. Using theoretical reasoning, extensive practical experience, and databases of prices the firm would identify pairs of assets the prices of which

\(^{17}\) A “swap” is a contract to exchange two income-streams, for example fixed-rate and floating-rate interest on the same notional principal sum.

\(^{18}\) See, for example, the list of LTCM’s major positions on August 21, 1998 given in Perold (1999, pp. C6-7).
ought to be closely related, which should over the long run converge (and in some cases had to do so), but which for contingent reasons had diverged: perhaps one was temporarily somewhat easier to trade than the other, and therefore more popular, or perhaps institutions had a particular need for one rather than the other. The fund would then buy the underpriced, less popular asset, and borrow and sell the overpriced, more popular asset (or take positions equivalent to these by use of derivatives, especially swaps). The close relation between the two assets would mean that general market changes such as a rise or fall in interest rates or in the stock market would affect the prices of each nearly equally, and long-run convergence between their prices would create a small but very low-risk profit for LTCM. By “levering” its own capital – in other words, performing arbitrage using borrowed funds and/or securities – LTCM could turn this small profit into a larger one; this also increased risk, but only to modest levels. The partnership knew perfectly well that over the short and medium term prices might diverge further, but the probabilities and the consequences of them doing so were carefully calculated by a statistical “value-at-risk” model, which measures the potential losses from adverse market movements (by the late 1990s such models were used by all the sophisticated institutional participants in the financial markets).

*Pace* standard accounts of LTCM, however, the firm did not simply assume that past price patterns would continue into the future, nor did it display an uncritical attitude to its risk model. Observed volatilities and
correlations were increased by explicitly judgement-based “safety factors” to take account of possible changes in markets and of possible deficiencies in the model. A consequence of this conservatism was that LTCM’s risk model predicted risk levels that were substantially higher than those actually experienced (until the 1998 crisis). The model predicted an annual volatility of net asset value of 14.5% while the actual volatility was 11%, and both figures were considerably less than the 20% volatility that investors in LTCM had been warned to expect. LTCM also “stress-tested” its trading positions to gauge the effect on them of extreme events not captured by standard statistical models or by recent historical experience, events such as the failure of European economic and monetary union or stock exchanges crashing by a third in a day. LTCM balanced its portfolio to minimize the consequences of such events, and sometimes purchased explicit insurance against their consequences. With a considerable presence in the Italian capital markets, for example, LTCM decided it was prudent to buy insurance against bond default by the government of Italy.

Was what LTCM did “arbitrage”? LTCM certainly described its activities as such, and in so doing it was simply adopting the standard financial-market usage of the term. However, LTCM needed to deploy some of its own capital as a foundation for borrowing, and its need to apply a value-at-risk model makes clear that its trading involved risk (albeit apparently modest risk), when in orthodox finance theory arbitrage is defined as being profitable trading
that demands no net capital and is entirely free from risk. An economist might therefore respond that a case-study of LTCM’s risky trading has no bearing on our understanding of arbitrage as a riskless theoretical mechanism.

However, amongst LTCM’s major positions (and one of the two most serious sources of loss) was a close real-world counterpart of the arbitrage that imposes Black-Scholes option pricing, the central theorem of the economics of derivatives. This position, taken on in 1997, responded to a price discrepancy developing in the market for stock index options with long expirations. As the name suggests, an index option is one in which the underlying “asset” is not the stock of an individual corporation but the level of a broad stockmarket index. Increasingly, banks and insurers in Europe and the U.S. were selling investors products with returns linked to gains in equity indices but also a guaranteed “floor” to losses. Long expiry options were attractive to the vendors of such products as a means of hedging their risk, but such options were in short supply. The price of an option is dependent upon predictions of the volatility of the underlying asset, and market expectations of that volatility (“implied volatility”) can be deduced from option prices using option pricing theory – indeed, that is one of the most important uses of the theory. In 1997, however, the demand for long-expiry options had pushed the volatilities implied by their prices to levels that seemed to bear little relation to the volatilities of the underlying indices. Five-year options on the S&P 500 index, for example, were selling at implied volatilities of 22% per annum and higher, when the volatility
of the index itself had for several years fluctuated between 10% and 13%, and
the implied volatilities of shorter-term options were also much less than 20%
per annum (Perold 1999, pp. A7-A8). LTCM therefore sold large quantities of
five-year index options, while hedging the risks involved with index futures
and sometimes short-expiry options. In effect, then, LTCM responded to the
discrepancy in prices by selling the options and “buying” and adjusting the
replicating portfolio. Complications such as the use of short-expiry options
aside, LTCM was conducting the arbitrage that option pricing theory posits.

The event that triggered LTCM’s crisis was the decision of the Russian
government on 17 August 1998 to default on rouble-denominated bonds and in
effect to devalue the rouble while permitting Russian banks temporarily not to
honour foreign-exchange forward contracts. LTCM had only a minor exposure
to Russia, but the precise form of Russia’s actions caused significant losses to
other hedge funds and western banks. A hedge fund called High Risk
Opportunities failed, and unfounded rumours began that Lehman Brothers, an
established investment bank, was also about to do so. Suddenly, market unease
turned into self-feeding fear. A “flight-to-quality” took place, as a host of
institutions sought to liquidate investments that were seen as difficult to sell
and potentially higher risk, and to replace them with lower risk, more liquid
alternatives. Because LTCM’s arbitrage generally involved holding the former,
and short selling the latter, the result was a substantial market movement
against the fund.
A similar flight to quality was, for example, triggered by the attacks of September 11, 2001. The key difference between the events of 1998 and 2001 is that the 1998 flight to quality was amplified, overlain, and in some instances contradicted by a much more specific process.\textsuperscript{19} LTCM’s very success had encouraged imitation: other hedge funds and many of the world’s leading banks, notably Wall Street investment banks, had either taken up similar arbitrage trading or devoted more capital to it. In aggregate, this body of arbitrageurs held positions broadly similar to those of LTCM, but some of them had greater exposure to Russia than LTCM had. To cover losses incurred there, they had to liquidate other positions similar to LTCM’s. As the prices of these moved against the arbitrageurs, they found themselves having to liquidate

\textsuperscript{19}In a minority of instances, LTCM (and, most likely, its imitators) held the \textit{more} liquid asset and was short the illiquid one. In Germany and France, for example, LTCM held (highly liquid) government bonds, hedged by paying fixed interest in (less liquid) interest-rate swaps. This kind of case is crucial in allowing the effects of a flight to quality to be distinguished analytically from those of the more specific process discussed in the test. Swap spreads (the difference between the fixed interest rate at which swaps can be entered into and the yield of government bonds of equivalent maturity denominated in the same currency) should rise in a flight to quality. They did indeed do so sharply in the U.S., U.K., and Sweden in 1998, but much less so in Japan (where LTCM had no net position), while in France and Germany they fell during much of the crisis. For further details, see MacKenzie (2003b).
further positions, thus further worsening price pressures, and so on. The arbitrage “superportfolio” (the aggregate of arbitrage positions similar to LTCM’s) began to unravel.

Paradoxically, the process seems to have been intensified by risk management practices in banks. Banks employ value-at-risk models not just as LTCM did (to gauge the overall risks faced by the fund), but also as a management tool (see, e.g., Dunbar 2000). By allocating value-at-risk limits to individual traders and trading desks, banks can prevent the accumulation of over-risky positions while giving traders flexibility within those limits. In 1996, the importance of value-at-risk models was increased when the Basle Committee on Banking Supervision permitted banks to use these models to help determine capital adequacy ratios. This reduced the amount of capital that banks had to set aside, but had the consequence that as volatility increased and prices moved against a bank, it faced a choice between setting aside more capital and liquidating its positions. In August 1998, many seem to have chosen the latter option. Value-at-risk became a stop-loss rule: the traders involved had no alternative but to try to cut their losses and sell, even if it was an extremely unfavourable time to do so. In August 1998, widespread efforts, apparently often driven by risk models, to liquidate broadly similar positions in roughly the same set of markets, in a situation in which those who might otherwise have bought such assets were also trying to sell, intensified the adverse market movements that were the initial problem. Crucially, these
various processes unravelling the arbitrage superportfolio led to greatly enhanced correlations between what historically had been only loosely related markets, across which risk had seemed to be reduced by diversification.

When used as management and capital adequacy tools, value-at-risk models (intended to describe the market as if it were an external thing) thus became part of a process that magnified adverse market movements, which reached levels far beyond those anticipated by these models. LTCM’s loss in August 1998 was a $-10.5\sigma$ event on the firm’s risk model, and a $-14\sigma$ event in terms of the actual previous price movements: both have probabilities that are vanishingly small. Value-at-risk models with stop-loss rules, other forms of stop-loss, management nervousness, fears by hedge fund managers of investor withdrawals, the need to liquidate positions to meet such withdrawals, cover losses and meet margin calls – all these seem to have combined to cause a failure of arbitrage. As “spreads” (the difference between prices of related assets) widened, and thus arbitrage opportunities grew more attractive, arbitrageurs did not move into the market, narrowing spreads and restoring “normality.” Instead, potential arbitrageurs continued to flee, widening spreads and intensifying the problems of those who remained, such as LTCM.

LTCM, however, was constructed so robustly that these problems, though they caused major losses, were not fatal. In September 1998, though, LTCM’s difficulties became public. On September 2, Meriwether sent a private
fax to LTCM’s investors describing the fund’s August results and seeking to raise further capital to exploit what (quite reasonably) he described as attractive arbitrage opportunities. The fax was posted almost immediately on the Internet and seems to have been read as evidence of desperation. The nervousness of the markets crystallized as fear of LTCM’s failure. Almost no-one could be persuaded to buy, at any reasonable price, an asset that LTCM was known or believed to hold, because of the concern that the markets were about to be saturated by a fire sale of the fund’s positions. In addition, LTCM’s counterparties – the banks and other institutions that had taken the other side of its trades – protected themselves as much as possible against LTCM’s failure by a mechanism that seems to have sealed the fund’s fate. LTCM had constructed its trades so that solid collateral, typically government bonds, moved backwards and forwards between it and its counterparties as market prices moved in favor of one or the other. Under normal circumstances, when market prices were unequivocal, it was an eminently sensible way of controlling risk. But in the fear-chilled, illiquid markets of September 1998, prices lost their character as external facts. LTCM’s counterparties marked against LTCM: that is, they chose prices that were prices that were predicated on LTCM’s failure. That minimized the consequences for their balance-sheets of LTCM’s failure by getting hold of as much of the firm’s collateral as possible, but made that failure inevitable by draining the firm of its remaining capital.

20 The fax is reproduced in Perold (1999, pp. D1-D3).
Consider, for example, the case of long-expiry index options, the source of nearly half of the September 1998 losses that took LTCM from late August’s less-than-fatal situation to the brink of bankruptcy. In what was becoming a global crisis, the volatility of the underlying indices rose, but to that general process was added a specific effect. As price quotations for the options that LTCM had sold rose (in other words, as their implied volatilities increased), LTCM had to transfer collateral into accounts held by its counterparty banks. Long-expiry options trading involves a limited number of people, most of whom are highly sophisticated. They knew this process was going on, and that if it (and analogous processes in the other markets within which LTCM operated) continued, the hedge fund would fail. If it did, LTCM’s counterparties would lose the hedge that LTCM had provided them, but they would at least then own the collateral that LTCM had had to transfer. So it was in the interest of each individual counterparty to “mark against” LTCM: in other words, to use price quotations unfavourable to the fund (see, e.g. Dunbar 2000, p. 213). The process would lead inevitably to the feared event – LTCM’s bankruptcy – but it would have been irrational for any individual bank to have sought to resist it by employing prices favourable to LTCM. If LTCM contested

---

21 Quantitative evidence consistent with this effect is that the rise of implied volatility of long-expiry options was much less in Japan, where LTCM had not been selling such options, than in the markets in which LTCM had sold large quantities of options (see MacKenzie 2003a, table 2, p. 368).
the unfavourable price quotations, as was its legal right, the dispute would have been settled by obtaining quotations from brokers not directly involved. Those brokers, however, would also have to take into account the effects of LTCM’s likely bankruptcy in offering their quotations.

LTCM’s failure thus became a “self-fulfilling prophecy,” in the sense of the phrase’s coiner, social theorist and sociologist of science Robert K. Merton, father of the finance theorist. The process was a little more complicated than his simple “sociological parable,” a run on a bank, in which investors fear that a bank will fail, seek to withdraw their deposits, and in so doing cause the bank to fail (Merton 1949). The structure of the process was, however, identical.

Discussion

It would be easy to conclude, as does most of the commentary on LTCM such as Dunbar (2000) and Lowenstein (2000), that the case shows that finance theory’s models are false: they do not correspond exactly to reality; they have assumptions that are not empirically correct. There is a sense, however, that while the conclusion is right, it is banal. All models are in this sense false, and are known to be false by their developers and by all sophisticated users of them; and this is the case for models in the natural sciences as well as in economics.
No model is, or ever could be, an exact replica of the reality modelled. For example, what took place in the long-expiry index options markets in 1998 was a process not envisaged in the Black-Scholes-Merton model, but that reminds us only that, rather than an attempt at exact replication of the empirical world, the latter is indeed a model as its developers always knew.22

What is needed, instead, is a much more nuanced understanding of the relationship between finance theory’s models and the financial markets. Arbitrage, this paper argues, is a useful topic to investigate from this viewpoint because of its pervasive role in financial models. This investigation obviously requires the perspective of the economist, but also those of the historian and sociologist.

Let me begin with economics. Part of the process that undermined LTCM’s arbitrage trading in 1998 was identified and modelled in a prescient article by Andrei Shleifer and Robert Vishny (1997; see also Shleifer 2000). The real-world practice of arbitrage, they point out, often involves risk and requires

22 For economically consequential ways in which the model and “reality” differ, see Black (1988).
capital. If those who invest in arbitrageurs (hedge fund investors, senior management in investment banks, etc.) are influenced by the performance of these investments, then a price movement against arbitrageurs can become self-reinforcing, as the latter are forced by investor withdrawals to abandon even excellent arbitrage opportunities. That is in essence what took place in 1998. There was nothing fundamentally wrong with LTCM’s portfolio. The spreads that the fund – and other arbitrageurs – expected to converge did indeed eventually do so: the consortium of banks that recapitalized LTCM not only recouped its capital but made a reasonable profit, and would have made a greater profit had its priority not been orderly but rapid liquidation of LTCM’s portfolio. Before converging, however, spreads diverged so radically that arbitrageurs incurred huge losses and had to flee the market.

In one sense, of course, this process was simply a manifestation of the dictum frequently attributed to J.M. Keynes (e.g. by Rubin and Weisberg 2003, p. 82): “Markets can remain irrational longer than you can remain solvent.” The key point emphasized by Shleifer and Vishny, however, is that the extent of this danger must surely vary from market to market and from one form of arbitrage to another. Some forms of price discrepancy are thus much less likely to be
eliminated by arbitrage than others, and so the fit between empirical prices and models is likely to vary considerably. Empirical, econometric investigation of “the limits of arbitrage” is growing – such investigation was surprisingly sparse in the past, but in 2002 an entire issue of the *Journal of Financial Economics* was devoted to it\(^{23}\) – and can be expected to provide considerable insights into the relative empirical success of different models.

Even the best contributions to this literature (e.g. Shleifer 2000) remain rather ahistorical. There is every reason, however, to expect the powers and limits of arbitrage to vary with time, and hence an historian’s perspective is also needed. In part, this is a matter of the history of technology and of economic history. Arbitrage is trading that exploits price discrepancies, and to perform arbitrage it is thus necessary to have up-to-date knowledge of prices. The availability of that knowledge – and the extent to which it is the property of a privileged few, or disseminated more widely – is clearly important. A classic discrepancy exploited by foreign exchange arbitrageurs was inconsistency between the exchange rates of different currencies, which could create

situations in which it was possible, for example, profitably to exchange sterling for francs in London while exchanging francs for sterling in Paris, or to exchange sterling for dollars, dollars for marks, and marks back to more than the initial sterling sum. The introduction in 1973 of Reuters’ electronic “Monitor” system, which disseminated exchange rate quotations worldwide to subscribers’ PDP8M minicomputers in close to real time, greatly facilitated arbitrage, and the traditional inconsistencies thus largely disappeared (Read 1999, pp. 363-66; Knorr Cetina and Bruegger 2002, pp.394-95). From 1975 to 1978 the prices of bonds, commodities and stocks were added to the service, again facilitating arbitrage.24

Much arbitrage is, however, more complex than exploitation of exchange rate inconsistencies and other simple discrepancies. This raises a subtler historical matter: the influence upon the practice of arbitrage of finance theory’s models, one aspect of Callon’s issue of performativity. When the Black-Scholes-Merton option pricing model was first published in 1973 the fit between the

24 Rubin and Weisberg (2003, p. 91) suggest a similar decline in discrepancies in gold and silver prices.
model and observed patterns of option pricing was only approximate (see MacKenzie and Millo 2003), and the use of the model in arbitrage was part of the process by which “theory” and “reality” converged. If this is typical – and it largely remains to be investigated – the historian of economic thought will need to join the economic historian and historian of technology for arbitrage fully to be understood historically.

The third perspective that is necessary is that of the sociologist. Part of the process leading to the crisis of 1998 was imitation. Imitation – “herd behaviour” it is sometimes called – in financial markets if often noted, but typically assumed to be the characteristic of naïve, uninformed, lay investors, and to be irrational. If the account presented here is correct, however, the preconditions of the 1998 crisis were created by skilled, professional investors imitating each other. The key, fatal, consequence of imitation was the disastrous increase in correlations of August and September 1998. LTCM’s arbitrage positions were geographically diverse, in disparate asset classes, and in spreads that at the level of economic “fundamentals” were often quite unrelated. Yet correlations that had historically been at the level of 0.1 or lower jumped during the crisis to around 0.7 as the holders of the imitative
superportfolio started (in a sense quite rationally) to liquidate its components simultaneously. This – the “social,” imitative, correlational nature of financial risk – is the lesson that LTCM’s principals have learned from the events of 1998. The risk model of LTCM’s successor fund, JWM Partners, incorporates the possibility that an extreme event can trigger all correlations to become 1.0.

The “sociological” risk of imitative correlation bears directly on the limits of arbitrage. Without it, arbitrageurs could manage Keynes’ risk – insolvency resulting from serious worsening of price discrepancies before convergence takes place – by diversification, as Shleifer and Vishny (1997) acknowledge. Arbitrageurs could, as LTCM did, take positions in a substantial number of “spreads” that were unrelated economically. If these positions could be guaranteed to remain uncorrelated, temporary losses in one or more positions could almost certainly readily be survived. If, however, crises induce correlations, for example because many others have also taken the same arbitrage positions, then diversification can fail to circumvent arbitrage’s limits, as it did in 1998.
Arbitrageurs, therefore, need to be studied from the sociologist’s viewpoint: as a community, not just as isolated individuals. Community does not of course imply harmony – as with the inhabitants of any village, deep hostility is amongst the possibilities – merely mutual susceptibility. Arbitrageurs in the same or related markets do in practice tend to know each other, they often talk to each other, and they have beliefs – sometimes accurate, sometimes not – about what each other is doing (see MacKenzie 2003). Because arbitrage is central to financial markets, their community is at the heart of high modernity, and deserves far greater sociological attention than it has received. Social processes within the arbitrage community may seem very distant from the abstract mathematics of Itô calculus and martingale theory. This article’s argument, however, is that the relationship of finance theory’s models to the markets those models describe cannot properly be understood without an economics of arbitrage, a history of arbitrage, and also a sociology of arbitrage.

25 There is a growing literature on the sociology of financial markets (see, e.g., Godechot 2001, and for wider references MacKenzie 2003 and MacKenzie and Millo 2003). Only a small part of this literature, however, focuses directly on arbitrage: see, especially, Beunza and Stark (2004).
REFERENCES


