European Societies
3(1) 2001: 69-90
© 2001
Taylor & Francis Ltd
ISSN
1461-6696 print
1469-8307 online

RELIABILITY AT RISK

The supervision of financial models as a case study for reflexive economic sociology

A. Javier Izquierdo
Universidad Nacional de Educación a Distancia, Madrid, Spain

ABSTRACT: This article focuses on financial 'model risk' supervision as a test case for a reflexive approach to the sociology of contemporary financial markets. Model risk is customarily defined as a statistically significant relation between the expectation of a trading loss by a firm and its strategic use of, somehow flawed, econometric models for asset pricing and market trading purposes. It made its first public appearance in the mid-1990s, as a component of a new generation of financial risk management systems for the financial derivatives industry. Since then it has been assimilated by the most sophisticated national and international financial regulatory bodies. A perfect illustration of the thesis of the progressive 'embedding of the economy into economics', the forensic practice of financial reliability trials (backtesting) faces a deep pragmatic dilemma: how to distinguish truly unpredictable error from negligent risk management behaviour in a wildly randomized social environment.
Key words: financial engineering; financial supervision; market risk; model risk; reflexive economic sociology; social randomness

> I believe that Value at Risk {econometrics} is the alibi bankers will give shareholders (and the bailing-out taxpayer) to show documented due diligence and will express that their blow-up came from *truly unforeseeable circumstances* and *events with low probability* – not from taking large risks they did not understand.
>
> (Taleb 1997b: 2)

> The kind of 'quants' who had thought up Value at Risk {econometric models} . . . were also the ones doing relative value and arbitrage trading. Was it possible that smart people could engineer their way round the failsafe mechanisms? Was it possible to fool VaR and take hidden risks?
>
> (Dunbar 2000: 147, on the 1998 débâcle of Long-Term Capital Management)

DOI: 10.1080/14616690120046950

1 Introduction

In an influential theoretical characterization of the radicalized, reflexive form of modern culture emergent in advanced industrial societies during the second half of the twentieth century, sociologist Anthony Giddens has stressed *technological reliability* as one of the key elements of this new ideal type of sociocultural configuration. The main distinctive feature of this new modality of 'impersonal trust' (Shapiro 1987) involved in modern institutions is precisely that it is 'vested not on individuals but on abstract capacities'. Giddens thus portrays *reliability* as a 'form of "faith" in which trust is vested on probable outcomes'. Being the product of a radical suspension of our common-sense judgement, the increasingly uncritical granting of the highest moral value to numerical measures of the degree of uncertainty of future events expresses 'a compromise with something more than mere cognitive understanding . . . {it represents in fact a reliance} upon vague and partial understandings of the "knowledge base" of expert systems' (Giddens 1993: 26–7). Among the most important sanctuaries where this new cult is preached, Giddens and the other champions of the theory of 'Second Modernity' invariably single out one particular social arena: the new, globally integrated variety of capital market.

More precisely, what these theorists frequently identify as the very epitome of an explicitly *reflexive* new form of social life are the complex sociotechnical assemblies of the sophisticated risk management systems used in the trading of financial derivative products and services such as futures, options and swaps (Giddens 1995: 153). Recent theoretical trends in the sociology of science and technology have developed a variation on the theme of technical reliability as impersonal trust-building around a network theory of 'heterogeneous alliances' which link the living and highly unstable memories of human biological bodies to the inert and durable memories incorporated in the natural and technical design of physical objects in the form of 'techno-economic networks' (Callon 1991). A natural sequel of the techno-economic networks approach to the sociology of financial markets has been the introduction of a generalized political model, namely that of advanced-liberal 'government at a distance' (Miller and Rose 1990). The global game of financial competition and regulation is then reinterpreted as a concrete realization of a more general trend in modern political culture to 'governmentalize' (Foucault) individual collective action by inscribing abstract expert knowledge, in the form of technical reliability, into the very institutional substrate of social arrangements (Porter 1995).

Having mainly dealt with the latter research programmes in my previous work on the sociology of financial markets (Izquierdo 1999a), I will here present a slightly different approach to the subject matter freely inspired by the reflexive sociological models of economic cognition and action developed by Mirowski (1990, 1991) and Boltanski and Thévenot (1991). My particular account of how the naughty phantom of social reflexivity terrifies contemporary financial competition and regulation will focus on one unexpected outcome of the huge industrial success attained by a social science endeavour: applied financial economics. After trying unsuccessfully to reduce to mathematics and computer logic a world of wildness nobody had constructed by design, financial economics has turned into an artificial social systems engineering project. This project, 'computational financial engineering', intends to build from scratch the kind of transparent cultural machines that could be fully understandable and controllable by the king-scientist. However, put to work in a market environment where stability is increasingly determined by their being accurate, computational (econometric) models of financial risk now have to face the perverse effects of their own success in the form of new, uncontrolled types of financial risk. Among these technologically induced, second-order types of financial risks is the so-called 'model risk', which has recently been the subject of intense regulatory controversy on an international scale.

The economics of financial modelling is discussed in the second section. The third section presents the concept of 'model risk', while the fourth offers the basic material on the regime shift in the domain of international financial risk supervisory procedures. The final section constructs a link between the uncertainty of conventional supervisory judgements on econometric models' performance and the ambiguous scientific status of the econometric theory of financial randomness. Some troubling sociological hypotheses are discussed in the conclusion.

2 The political economy of financial risk modelling

We can distinguish between two separate *economic* uses *of financial risk models*. First, there is what may be called an *industrial use*, associated with the cost and long-term monopolistic returns of the competitive strategies devised by individual firms in the incipient marketplace for computational financial risk management systems. In the short run, however, the main preoccupation of the users of this financial experts system is the direct *financial use* of internal risk control models: the gains in allocative efficiency of capital reserves obtained as a consequence of disposing of more accurate mathematical models for financial risk management.

The supervisory controversy over sound bank internal risk management systems and safe capital reserves (Swary and Topf 1993) falls into this second economic dimension of financial econometrics expert knowledge.

The level of capital safety requirements of financial intermediaries (the ratio of reserves to financial assets) is a key factor in the market competitiveness of these firms. Traditional base asset–liability figures in the balance sheets of financial dealers need to be complemented with new types of standardized risk-accounting data that change the level of capital safety requirements. Due to the new supervisory regulations (see below), the quantitative calculus of expected trading losses internally performed by a financial firm has a direct and strong effect on the level of reserves required to fully insure a firm's creditors and shareholders against bankruptcy. Thus it has a direct effect on the firm's financial bottom line and profitability.

However, the human, social activity of mathematical economic and econometric modelling is still amenable to a third, more direct and explicit kind of economic analysis in terms of cost–benefit and risk–return calculations: a second-order type of financial risk, known by financial analysts, engineers and traders as *model risk* (Derman 1996a, 1996b). The publicized débâcle, in September 1998, of the large and sophisticated hedge fund Long-Term Capital Management is the most telling example of the devastating effect that can be produced by these strange forms of financial risk: a truly *reflexive* form of economic risk that is produced by the actions of risk-adverse financial agents themselves, using mathematical asset-pricing models in an intensive and extensive manner to build financial insurance policies or risk-hedging instruments: the famous 'financial derivatives' products, such as futures, options and swaps contracts (Steinherr 1998).

3 Defining model risk

Model risk has been defined as a kind of financial risk that 'results from the inappropriate specification of a theoretical model or the use of an appropriate model but in an inadequate framework or for the wrong purpose' (Gibson *et al.* 1998: 5). The particular risks and uncertainties implied by the practice of formal scientific inquiry (modelling, estimating and testing) into the economics of financial markets activity are the very 'fundamentals', in the economic sense, of the market value of formal financial knowledge, understood as a key competitive resource in the modern world of finance. Hence, the multiple *economic sources* of *model risk* are associated with the almost infinite manners of constructing a wrong theoretical model or using a correct model in the wrong way.

In a first, economic approximation, the concept of *model risk* accounts for the fact that the existence and utilization of different types of formal asset-pricing econometric models can give rise to a wide diversity of theoretical prices for a similar type of financial product. As the discrepancy between these theoretical bid–ask prices resolves itself in the market process, the use of theoretical prices as inputs for the decision-making process of trading and dealing in real financial markets is revealed as a major factor of economic success and failure in contemporary financial global competition.[1] The second, methodological approach to the concept of *model risk* focuses on the existence of different types and levels of error in the practice of economic modelling: at the base-theoretical hypothesis, translation into mathematical expression, statistical data inputs, arithmetical calculations, computer softwiring or trading misuses.

Model risk in financial markets appears each time an asset-pricing model does not take into account some relevant factor of price variation, or else wrongly assumes that the motion of certain stochastic variables can be imitated by a deterministic process, or thinks that price changes can be described by a normal frequency distribution with limit variance range. In other cases – even if the model could be thought of as 'correct in principle', or at least not patently erroneous from the point of view of the formal logical arguments, mathematical proofs, probabilistic test and 'encompassing' checks commonly used in academic econometric diagnostics – markets can disagree with its results in the short term. The data used can also have been badly estimated or collected, or there may have been a mistake during the heuristics searching for its analytical solution. The model may also have been badly calibrated to mimic real market statistics. There may even have been coding errors in programming it into the computer, or the model may have been used in an incorrect way by the final user (e.g. a trader may have applied it to price for instruments or markets for which it lacked predictive validity), and so on.

As has been observed by most finance scholars and professional derivatives traders, the core mathematical and statistical assumptions built into standard neoclassical pricing models suffer tremendously when they confront the structures and processes of real-world financial trading and firm-wide risk management. While equilibrium asset-pricing models, for example, characteristically assume that markets are composed of atomized agents who cannot substantially influence each other or individually manipulate aggregate market prices, imitative contagion and herd behaviour are ubiquitous in real markets and giant, reputed investors occasionally also 'move the markets'. Common financial models furthermore take

1. A New York-based financial consultancy, Capital Market Risk Advisors (CMRA), recently estimated that 40 per cent of total derivatives-related trading losses were due to modelling errors: up to 2.7 billion US dollars during 1997. This same consulting firm estimates that in the period 1987 to 1997 some 4.7 billion of total cumulative derivative losses of 23.8 billion dollars were due to pricing errors caused by wrong data or wrong assumptions in asset-pricing models (Stix 1998: 27).

for granted that economic information is a public good, while in practice there are different rhythms of accessing and analysing it. They also assume that transaction costs are minimal and that markets are highly liquid; but liquidity 'squeezes', and large jumps in prices and volatilities are all typical of real, legally and organizationally constructed markets. The same applies to the standard assumption of levelled debt capacity and regulatory neutrality. There is a wide variation in the financial and legal costs of running a banking business depending on the different institutional and social statuses of the agents.

4 Supervising model risk: technical controversies and public choices

Byzantine academic debates over how to define, measure and reduce model risk are central to the supervisory controversy over the calculation of so-called 'market risk' banking capital requirements. Having proved powerless to accommodate its standard bureaucratic norms for external banking examination to the ever faster rhythm of technical innovation in financial derivatives markets, the main international banking supervisory agency, the Basle Committee for Banking Supervision (BCBS) of the Bank of International Settlement (BIS), has recently given a Copernican-turn to the tradition of central banking supervision, a tradition whose most conspicuous example is the 1988 Basle Capital Accord (BCA) (Swary and Topf 1993: 133–4). Confronted with the constant failures of mandatory and universal supervisory standards, the BCBS now tries to enlist into its team the adaptive powers of the decentralized mechanism of innovation-based market competition that allows most financial firms to continually improve internal risk management systems by heavily investing in human capital and R&D (Dunbar 1998).

Setting global market risk supervisory standards

The BCBS intended to integrate the fast-evolving organizational know-how of the derivatives industry into its extended supervisory repertoire – the 1996 Amendment to the BCA (ABCA) – by targeting the bank's own internal control systems and not, as was previously done, its real investment portfolio. At the end of the 1980s, the trading book and off-balance operations (mostly derivatives contracts) had gained so much space in the balance sheets of the savings and loans and commercial banks that the national and international regulatory authorities began to fear that, together with traditional credit risk, retail banks would now be strongly affected by that class of devastating risk specific to the investment banking and securities dealer business, namely, *market risk*. Authorities perceived an increasing probability that an adverse, sudden and coordinate price movement across diverse markets, terms and instruments worldwide could produce such a huge quantity of trading losses that the precautionary capital reserves, which serve as guarantees for depositors, would be severely affected and trigger a spiral of financial panics and bankruptcies. With the US savings and loans disaster reaching its peak at the beginning of the 1990s, the initial rhetorical concern of public authorities over market risk translated into a concrete programme for adapting regulatory capital requirements to the new reality.

In 1988 the BCBS succeeded in having its members sign the first international protocol for harmonizing national banking capital standards: the Basle Capital Accord (BCA). The BCA prescribed the acceptance of a set of common procedural rules, a system of direct external supervision known as the 'standard approach' (Basle Committee 1988/1998). By mechanically applying the same broad criteria for credit analysis, the different national authorities could determine, in a crude but normalized way, what should be the correct and safe level of capital reserves for a bank in possession of a diversified credit portfolio to insure its depositors and shareholders against a huge wave of credit defaults, regardless of the national legislation. This common measure of banking safety was known as the 'Cooke Ratio'.[2]

However, only two years later the supervisory norms of the BCA had became outdated by the new investment practices of its regulatory subjects, that is, by massive exchange-traded and OTC[3] derivatives trading. The BCA strictly focused on the regulation of *credit risk capital requirements*, the amount of capital that must be set aside to insure banks' bottom lines against risks of credit default, and said almost nothing about the incipient problem of market risk precautionary capital.

Thus, shortly after the BCA began to be applied by national authorities, the BCBS was already seriously entertaining the possibility of amending it and including new precautionary standards against market risk. A

2. The BCA required banks to raise their reserve to reach at least 8 per cent of total assets, weighted by risk class. It distinguished two components or 'tiers' of banking capital: Tier 1 or 'core' capital (stock issues and disclosed reserves), and Tier 2 or 'supplementary' capital (perpetual securities, undisclosed reserves, subordinated debt with maturity greater than five years, and shares redeemable at the option of the issuer). Finally the Accord established a set of risk capital weights to ponder capital requirements against different types of financial instruments (Swary and Topf 1993: 450–6).
3. OTC is for 'over-the-counter' or tailor-made derivatives contracts, such as foreign exchange options or so-called 'swaptions' (options on interest rate swaps). Contrary to publicly exchanged financial securities OTC derivatives are privately negotiated mainly between an investment bank and its client corporation.

new regulatory proposal was devised to encourage the international adoption of a new simple, transparent and amply agreed procedure to determine with sufficient precision the extra quantity of capital reserves needed by the banks with huge portfolios of derivatives and other high-risk securities.

At the end of 1996 the BCBS issued an advisory report that recommended banks to use their own *internal risk measurement models* and their own computerized systems of firm-wide risk management to determine for themselves the proper quantity of market risk capital reserves (Basle Committee 1996a: 38–50). There was a double argument in support of this proposal: (1) to profit socially from the private information and entrepreneurial know-how accumulated during years of daily risk management, and (2) to publicly profit from the firms' own selfish interests in improving the quality of its risk management system to gain competitive advantage. With the coming into force in January 1997 of the Amendment to the BCA (ABCA) that allowed banks to use their own internal risk management models to autonomously determine the proper amount of market risk capital reserves, public supervisory authorities have come to perform rather indirect and abstract new inspectorate tasks, centred around a set of very technical procedures for risk management systems quality auditing.

In this new regulatory regime, effective banking safety levels can only be guessed indirectly by supervisory authorities, by means of checking the technical reliability and organizational flexibility of banks' internal risk management systems.

The design of banks' internal control systems: value-at-risk econometric modelling

Opposed to the former 'standard approach' to banking supervision, the new supervisory regime for market risk capital reserves is known as the 'internal models approach' (Jorion 1997a: 50). Many of the internal risk control systems developed by the banks who are active in the global derivatives markets are based on the application of a class of generalized equilibrium asset-pricing econometric models known as Value-at-Risk (VaR) models. The basic principle of VaR management, the daily calculation of a broad, aggregate figure of maximum potential losses, had been developed within the community of the biggest Wall Street investment banks almost since the aftermath of the October 1987 stock-market crash.

VaR models tackle the following computational problem: how to determine the maximum financial loss, expected with a significant probability for a given confidence level, that could be suffered by a properly diversified asset portfolio during a given period of time, as a consequence of an adverse and pronounced movement in financial prices coordinated across different markets, instruments, maturities or countries (see Jorion 1997a: 86–93). Technically, a VaR figure is a probabilistic measure of future economic value, or, to be more precise, a mathematical expectation of financial losses defined as the mean probability associated with a given event times the economic value assigned to this event. The information provided by VaR numbers is an estimation of the maximum pecuniary losses (e.g. five million euros) attached to a numerical probability of occurrence (1 per cent), a statistical confidence level (99 per cent) – and therefore to some theoretical frequency distribution (e.g. gaussian) – and a period of time (one day). That is, of each 100 trading days one should expect that only during one of these one's investment portfolio could reach a maximum cumulated daily loss of five million euros, and that with a margin of error of ±1. The amplitude of this error interval thus accounts for the possibility of a maximum-loss event occurring twice during the chosen time period.

The most common procedure used to calculate VaR figures is called the 'historical method'. This is a two-step econometric procedure originally codified by JP Morgan into its proprietary risk management software *Riskmetrics*™ (Guldimann 2000). It works in the following manner. It is first of all necessary to arrange a complete and extended numerical database, that is, a multidimensional matrix of previous fundamental parameter changes in the most frequently traded financial instruments. This should constitute a reliable sample of the long-term behaviour of markets and will allow the user to estimate a set of robust statistical trends in the relations between (1) the *market prices* of a broad range of investment contracts (end-of-the-day quotes of shares, index, bonds, futures, etc.), (2) its *volatilities*, that is, the mean deviations of every single market price from its mean historical level; and (3) its *correlations*, or the statistically significant coefficients of mutual influence between the long-term motion of each security and the historical motion of each and every other security related to it. These three types of sample statistics (mean values, volatilities and correlations) are the variables which are subject to econometric treatment within VaR models, typically constructed in the form of equilibrium asset-pricing models obeying the well-known mean–variance principle of neoclassical finance theory (optimal risk spread, defined as the minimum aggregate variance of mean expected returns for any given level of subjective risk-aversion).

A much used alternative approach to VaR calculations – and favoured by Bankers Trust with its computer application *RaRoc2020*™ (Falloon 1995) – is taken not from classical portfolio theory but from the theory of arbitrage-free option pricing (Jorion 1997a: 77). In this case the key variables of the model are not correlations or historical volatilities but fundamental risk parameters that can be derived from the Merton–Black–Scholes option-pricing model: *delta*, *gamma*, *vega*, *theta*, *rho*, etc. In this approach each financial contract is decomposed or 'granulized' into a series of basic risk factors: 'delta-risk', 'gamma-risk', etc. (Merton 1995a). Huge masses of these little risk 'grains' or 'particles' are then aggregated using statistical correlation techniques, until a single figure results that measures the risk-adjusted return on all the capital invested in the market. Two other statistical simulation techniques are widely used to complement the analysis in terms of historical volatilities and risk factors: Monte Carlo simulations (based on artificially calibrated computational samples and stochastic processes) and 'stress testing', a qualitative assessment of the robustness of different portfolio structures under extreme-value conditions (see Dunbar 1999).

Reliability trials: backtesting

The 1996 ABCA established a series of minimum general 'technical' requirements that banks' internal risk management systems need to fulfil. The initial validation and periodic revision of bank internal models under its jurisdiction was a task assigned to national banking supervisory authorities. The amendment of 1996 was also accompanied by a complementary advisory report that established a set of criteria for national supervisory authorities to conduct quality audits of banks' VaR internal models (Basle Committee 1996b). The aim of this complementary report on 'backtesting' procedures was to add an incentive mechanism for compliance with regulatory norms to assure the public that if banks wanted to gain supervisory approval for using their internal risk management systems as 'regulatory allies', they would have to adopt the necessary (and costly) measures to improve their accuracy.

The report in question detailed how to conduct a series of standard statistical counter-trials or 'backtests' to formally asses the performance of bank internal models' risk measures in relation to the actual risk levels in the market. To guarantee that banks would indeed devote the required efforts and resources to maintain, update and improve their internal models, the report stipulated that the different national supervisory authorities would conduct quarterly examinations of their forecasting performance. These exams would monitor the quality of the internal statistical information used by bank CEOs in the decision-making process to set a safe level of market risk capital reserves. Hence, the ultimate aim of the model examination is to guarantee that the VaR figures of aggregate financial risk would comply with some minimum econometric reliability requirements.

As defined in this 1996 BCBS supplementary document, backtesting trials consist in the comparison of VaR *theoretical* measures calculated by a particular financial econometric model for a time horizon of one day, with *actual* financial profit and loss daily figures, that is, the effective 'trading outcomes', realized at the end of each business session (Basle Committee 1996b: 2). As we have seen, theoretical VaR measures are intended to encompass within them (almost) all trading outcomes expected at the end of the day, leaving outside of its coverage only a tiny fraction of these (i.e. the most improbable ones), whose size is given by the confidence level chosen to calibrate the model. In this respect the BCBS report established that the percentage of trading outcomes that the theoretical VaR measures produced by the banks must cover should be 'consistent' with a confidence level of 99 per cent.

Therefore, to assess the degree of statistical effectiveness of a bank's VaR econometric models, the public examiner must (1) count the number of 'exceptions' produced by the model, that is, how many times the actual trading outcomes at the end of the day *fall outside* the theoretical expectation produced by the model; and (2) determine if the number of exceptions is consistent with the obligatory coverage level of 99 per cent. For example, for a recommended sample of 250 trading days, a daily VaR measure calibrated for a 99 per cent confidence level should cover, on average, 248 of the 250 observed trading outcomes, leaving only two exceptions unforecasted by the safety calculus.[4] If the model produces, say, 125 exceptions, it must be 'clear' to the external public auditors that something is wrong. The bank must then compensate for the forecasting weakness of its model with a proportional rise in the multiplying factor applied to its capital reserves that happens to attain the desired confidence level of 99 per cent.

However, the main problem with which VaR econometric models external examiners have to deal is how to interpret an ambiguous backtesting result. That is, still using the former example, one that produces a number of exceptions only slightly higher than two – say four or seven – a figure that, from a strictly probabilistic point of view, is not a conclusive signal about the actual predictive strength or weakness of the model. To solve this fundamental supervisory uncertainty the BCBS document established a second set of quantitative criteria to clearly demarcate three different *interpretative zones*: a 'safety' zone (green), a 'caution' zone (yellow) and a 'danger' zone (red). The green zone extends to all backtesting results – between zero and four exceptions in a normalized sample of 250 – that

4. To make a trade-off between the regularity of the supervisory exams and the representativeness (in the statistical sense) of the data used by the models, the BCBS recommended carrying this backtesting exam on a quarterly basis, the evaluation focusing on trading data from the last twelve months, i.e. a sample of 250 observations.

'from a {mathematical} probabilistic point of view' suggest no doubts about the predictive soundness of the model. In this case no supervisory action is undertaken in the sense of rising capital requirements. Within the yellow zone fall those results that produce non-conclusive doubts about the forecasting ability of the model – between five and nine exceptions – and whose reading by the supervisor could be accompanied by a rise of between 0.40 and 0.85 points in the multiplying factor applied to the existing base capital reserves. Finally, those outcomes which are equal to or exceed 10 exceptions are located in the red zones, and all must be countered by a one-point rise in the multiplying factor.

Again, this system of zones has its own problems, as the supervisory report recognized. If the examiner is too stern about the numerical thresholds that demarcate the different zones she can commit two types of statistical errors in her lecture of backtesting results: either she can classify as defective a model that is actually valid, or she can admit as correct a model that is actually faulty. These types of problems are largely posed by those backtesting results which are included within the yellow zone, because standard statistical calculations show that the probabilities for a model to produce outcomes between five and nine exceptions are similar for acceptable (99 per cent coverage) and rejectable (98 or 97 per cent) models.

To aid the examiner to overcome this problem, the BCBS report included two tables with numerical calculations of existing theoretical probabilities to obtain a given number of exceptions for a sample of 250 observations for different coverage levels of the model (99 per cent, 98 per cent, 97 per cent, 96 per cent and 95 per cent). These calculations show that there exists a high probability of erroneously rejecting a valid model when, for a confidence level of 99 per cent, the examiner chooses a particularly low number of exceptions as the threshold for rejection (if the threshold is set to one exception, valid models would be rejected by examiners in 91.9 per cent of cases). Of course, if the threshold of the maximum number of exceptions that can be produced by a model to be validated is raised, the probability of incurring this type of error is lowered. However, the probability of making the inverse error is raised: for a rejection threshold of seven or more exceptions, the calculations of the Committee indicate that a model with a coverage of only 97 per cent (a non-valid model) will be erroneously accepted in 37.5 per cent of cases.

5 Types of randomness, error and responsibility

A further answer to the problems posed by of the ambiguity of backtesting results is provided by another Basle Committee recommendation.

The Committee eventually advises the supervisor to require the bank to supply a set of complementary information of a qualitative nature, both about the precise econometric and computational architecture of the model under supervision and about the 'special' character of non-covered trading outcomes.[5] This means that when there is not enough quantitative evidence about the technical reliability of the risk model, banks are still allowed to try to document, explain away and possibly justify, on a case-by-case basis, the causes of every exception detected through the backtesting.

The bank's model risk counter-experts do in fact routinely elaborate complex interpretative documents to try to explain away even the most flagrant backtesting exceptions. If, for example, a bank were to fail to raise its bottom-line capital level to insure creditors against adverse asset price movements produced by an abrupt social rupture in a foreign country, the bank VaR modellers would present supervisory authorities with newspaper clips and dossiers that qualify such an exceptional 'exception' as one of those completely unpredictable and hence uninsurable random economic events that supervisors conventionally allocate to the correct probabilistic margin of 1 per cent normal measurement error.[6] However, if the same failure were to apply to the occurrence of an adverse price change of the kind that is considered by neoclassical financial economists to be strictly governed by so-called 'endogenous market forces', such as recurrent stationary cycles in aggregate consumer demand or stable stochastic trends in macroeconomic growth rates, the fact of an eventual bankruptcy could hardly be publicly justified as the consequence of unnoticed and

5. 'The burden of proof in these situations should not be on the supervisor to prove that a problem exists, but rather should be on the bank to prove that their model is fundamentally sound. In such a situation, there are many different types of additional information that might be relevant to an assessment of the bank's model' (Basle Committee 1996b: 8).
6. The tale of the 'perfect financial storm' is *grosso modo* the scheme of the justificatory arguments put forward by defendants in the governmental inquiry that was set up after the private bail-out of the large hedge fund Long-Term Capital Management going 'technically bankrupt' in September 1998. In this particular account, the star role of the 'extreme event' is played by the default of Russian sovereigns (Dunbar 2000: xiii). Curiously enough, the fact of not being directly subject to Basle Committee internal models' regulations was one of the reasons for the fund's extraordinary success as 'global central banker for volatility' during the aftermath of the autumn 1997 Asian crisis (ibid.: 178), but also played an important role in its eventual debacle exactly one year later. In his careful reconstruction of the LTCM catastrophe, financial journalist Nicholas Dunbar claims that, despite the shock of the Russian bonds default, the real problems of the fund were in a larger part caused by the growing management prominence conceded to 'Risk Aggregator', the flawed, in-house VaR management software of LTCM. 'The Risk Aggregator has been the subject of much debate. As is now clear, it either didn't work properly or was misused by the LTCM partners – none of whom will now accept responsibility' (ibid.: 186).

unintended ‘modelling errors’ in the face of ‘radical market uncertainty’. The surest bet here for the supervisory examiners should be the presence of strategic ‘fake’ movements intended to make cheap, low-quality financial risk management policy appear to comply with high-quality, high-cost risk management supervisory standards. What I would like to suggest here is that serious doubts and criticisms from academics and practitioners alike have recently crept into this regime of conventional, peaceful techno-economic coordination between private bank modellers and supervisory examiners. To get rid of the frightening ghost of sudden financial débâcle no longer suffices to magically conjure, as do conventional financial modellers, the perfect isolation of stable economic functions from non-stationary sociohistorical processes.

Adopting the language of ‘standard econometrics’ as common currency in the political debate over global financial stability is no longer as unconscious an administrative behaviour as it used to be. To be sure, the mid-1990s academic controversy over the management and regulatory uses of VaR econometric models has produced a large repertoire of methodological, theoretical and epistemological justifications for adversarial types of econometric practice.[7] Among the most remarkable arguments put forward in this detective-forger social reflexive game is the banks’ risk modellers accusation of *arbitrariness* formulated against public supervisors for setting the standard confidence levels according to which backtesting results are to be judged in complete disagreement with the empirical statistical structure of real market fluctuations. When you choose a confidence level of 99 per cent it means that only one out of each 100 trading days your losses can exceed the VaR value computed by the model. But the true meaning of the confidence level is really an artefact of the adoption of a more fundamental (and disputed) theoretical assumption, namely that of a characteristic probability distribution. In neoclassical financial econometrics, statistical confidence is but the offspring of gaussian mathematical laws (the well-known ‘ergodic’ and ‘central-limit’ theorems), and when these mathematical theorems are rejected as a proper algorithmic

7. A fast foray into this controversy is provided by the published exchange between two financial experts, Philippe Jorion, finance professor at the University of California, Irvine and one of the principal academic advocates of VaR models, and Nassim Taleb, a respected senior option trader and derivatives engineer who is critical of VaR (see Jorion 1997b; Taleb 1997a, 1997b; Stix 1998). For Jorion, on the one hand, the purpose of VaR models is not, as is usually stated, ‘to describe the worst possible outcomes’ but, more modestly, ‘to provide an *estimate* of the range of possible gains and losses. Many derivatives disasters have occurred because senior management did not inquire about the first-order magnitude of the bets being taken’ (Jorion 1997b: 1). Taleb, on the other hand, discredits VaR econometrics as mere ‘charlatanism’, arguing that ‘it tries to estimate something that is not scientifically possible to estimate, namely the risks of rare events. It gives people misleading precision that could lead to the buildup of positions by hedgers. It lulls people to sleep’ (Taleb 1997a: 1).

representation of the empirical frequency distribution of price changes so is statistical confidence as a means for technological reliability.

Following the path initially tracked by the same financial firms they audit, supervisors have a decidedly 'mild' conception of financial randomness. But, as has been pointed out many times by the most incisive critics of financial neoclassical econometrics, there exists a flagrant gap between the tractable mathematical models of mild randomness generally assumed by applied portfolio theory and the type of 'wild' randomness in which, as is characteristic of true *historical processes*, extraordinary events are always, in some sense, 'too probable' (Mandelbrot 1997b: 57–74). Still, public regulators and private financial competitors alike have traditionally preferred to assume that 'randomness' is the source of mostly insignificant and easily reversible economic events; and that truly irreversible economic events, such as large-scale or long-term price variations, have nothing to do with randomness but are the product of deterministic, necessary and thus predictable causes.

This classical, reassuring principle for the administrative vision and division of the world – the well-known gaussian axiom that randomness can only be understood as a microscopic phenomena – is today in trouble in the world of derivatives trading. As much by the sheer brutality of recent market events as by the strategic necessity to adapt to changes in public supervisory norms, financial practitioners have been called upon to reflect upon the obscure and disputable modelling conventions that sustain the myth of technological *reliability* in the world of applied financial econometrics. In fact, even the very senior executives who run the risk management divisions of the biggest world investment banks are beginning to doubt the key feature of neoclassical financial theory and engineering practice: that you can separate deterministic from random forces.[8]

The irony here is that the strong point put forward by rational (scientific) criticism of financial management and regulatory practice is in this

8. Witness the crystal-clear account by prominent market professional, Robert Gummerlock, former managing director of Swiss Bank Corporation, one of the world's biggest investment banks: 'The magnitude of a 5–10 standard-deviation move is not debatable – that is given. What is debatable is how often it happens and that's where people get confused. In the textbook world of normal distributions, a 10 standard-deviation move is more than a one in a million event. In financial markets we know it is not; so we have to decide how often it can happen. The troublesome thing about fat tail distributions is that they sever the link between ordinary and extraordinary events. Under a purely normal distribution, the extraordinary events are strictly governed by probabilities, policed by the standard deviation. With fat tailed distributions, outliers can occur with maddening frequency and no amount of analysis of the standard deviations can yield useful information about them' (cited in Chew 1994: 64). It is indeed remarkable that practitioners' indictments against orthodox statistical financial risk measurement do read almost exactly the same as some of the most recent public statements by the very nemesis of academic neoclassical financial econometrics. 'The

case, and as it should be, *totally unacceptable* for supervisors. The reason for this is that, to accept the statistical spectre of 'wild' randomness as a more accurate scientific description of the typical spectral shape of real-word financial risk would mean to reject any role whatsoever for public supervision in the financial services industry.[9]

Minimum supervisory requirements for banking capital reserves only make sense in a world were financial risk is *statistically deterministic*: it can be modelled as a predictable phenomenon in the probabilistic sense, and therefore as something that falls under the domain of human control, even if this control is exercised under the subtle mathematical routines of stochastic dynamic programming (Sent 1998). For banking capital risk supervision to have a positive social welfare effect, financial catastrophe must be understood as something that *can be prevented*. For only under this hypothesis can some level of regulatory capital reserves be called *safe*, or a sudden bankruptcy attributed to *a failure to comply* with supervisory requirements. Using this 'classical' framework of analysis, financial management can be judged to have 'failed' and legal responsibility for 'mismanagement' can be sought on an individual basis.

However, if the speculative motion of financial prices is a non-deterministic process of a second-order class, as critics of neoclassical financial econometrics argue, then financial catastrophe cannot be privately or socially prevented. In this later scenario, no regulatory level of risk capital reserves (including full investments coverage) can be really deemed 'protective', and no financial damage to the bank's creditors or shareholders (even instantaneous bankruptcy) can be understood as the product of 'mismanagement'. Human responsibility is rather translated into the language of unforeseen, unintended random 'error'. In this

mathematics underlying portfolio theory handles extreme situations with benign neglect: it regards large market shifts as too unlikely to matter or as impossible to take into account. . . . According to portfolio theory, the probability of these large fluctuations would be a few millionths of a millionth of a millionth of a millionth. (The fluctuations are greater than 10 standard deviations.) But in fact, one observes spikes on a regular basis – as often as every month – and their probability amounts to a few hundredths' (Mandelbrot 1999: 70).

9. But also, paradoxically, to deny any productive role for the financial engineer's computational stylization of the economic process! As has been acknowledged by Peter L. Bernstein in his bestseller history of the triumphal march of mathematical financial economics in the academy and the marketplace, 'Mandelbrot remains on the periphery of financial theory, *both because of the inconvenience to analysts of accepting his arguments* and because of *the natural human desire to hope that fluctuations will remain within familiar bounds*' (Bernstein 1992: 132; my italics added). The said Benoît Mandelbrot has recently restated his old arguments as to the weak scientific status of financial econometrics, taking financial engineering as a new target for his clever invectives. 'Avant de s'engager dans l'ingéniérie financiere et ses "produits derivés", il s'impose d'abord de "s'assurer bien du fait" . . . on ne laisse pas à l'ingénieur le loisir de prendre à sa charge les regrets du savant' (Mandelbrot 1997b: 9).

alternative theoretical framework, it should come as no surprise that bank managers' overall judgement on public supervisory procedural norms for conducting model risk audits is that they are doing more harm than good to our collective economic welfare.

Tearing down the conventional administrative boundaries that separate ordinary from extraordinary economic events, as suggested by the Mandelbrotian hypothesis of 'wild' randomness, would imply that those management decisions, backed up by VaR results, concerning precautionary capital allocation that were considered the most flagrantly 'unjustifiable' under the gaussian, statistically deterministic supervisory framework, could be excused as the product of 'sheer bad luck'. Thus, confronted as it is by the hyperbolic stochastic dynamics of contemporary financial prices, the everyday administrative banking maintenance of the twin social constructs that define the institutional core of a capitalist market economy – *accounting value* and *commodity money* – desperately demands 'that something be treated as effectively invariant, even as we know all along it is not' (Mirowski 1991: 579).

6 Conclusion: forgers and critics

An intellectual adventure barely forty years old, the mathematical economic theory of equilibrium asset pricing in perfectly competitive capital markets has by now consolidated into one of the most dynamic and respected subfields of economics. Together with the exploding job market and 'indecent' salaries paid to hundreds of young MBAs in mathematical finance during the past two decades, the 1991 and 1997 Nobel prizes in economics awarded to the pioneering models of 'efficient' portfolio selection, risk pricing and capital arbitrage economic routines by Harry Markowitz, Williams Sharpe and Merton Miller, and the 'optimal' dynamic risk management and synthetic (derivative) asset replication schemes by Robert C. Merton and Myron Scholes stand as irrefutable proof of the scientific, economic and political success of this most esoteric body of social knowledge. On the other hand, highly publicized recent derivatives-driven financial catastrophes, such as the October 1987 NYSE market crash (Jacobs 1999), Metallgessechschaft and Orange Country in 1994 (Jorion 1995), Barings in 1995 (Millman 1995) and Long-Term Capital Management in 1998 (Dunbar 2000), have raised serious concerns about the scientific shortcomings and technological dangers of applied mathematical financial economics, so-called *financial engineering*.

In our contemporary high-tech world, the ancient dialectics of public domain expertise versus private and secret information has reached a peak in the arena of global financial markets' competition. The complex, reflexive social patterns characteristic of the world of derivative financial products, services and markets engineering – with its characteristic sequence of innovation, competition and regulation cycles (Abolafia 1996) – offer one of the most intricate present variants of this classical dialectics of authorized knowledge corrupted into strategic forgery and then recycled as learned criticism.[10] This new technological art of the artificial 'replication' or 'computational synthesis' of historically stable economic functions (Crane *et al.* 1995) is being increasingly self-consciously understood by its practitioners as a sort of *counterfeiting game*.[11] In such a game of economic competition social reflexivity is pervasive and, sooner or later, winning strategies are defeated as a consequence of their own success. No matter how impressive its trading record in the short term, the 'risk fake' manufactured by the financial engineer would eventually have to be authenticated by the most harsh critic of economic technology, namely economic history (Mandelbrot 1997a: 17–22).

In a series of working papers and official advisory reports published during the second half of the 1990s, the Basle Committee established *de facto* supervisory procedural rules for the correct way to conduct standard forensic trials to test the technological reliability of banks' risk control models. Avant-garde academics and professionals have lately attacked standard backtesting model risk supervisory methods for being unable to acknowledge the probabilistic subtleties of real-world financial risk. This particular denouncement reproduces and updates the well-known market-libertarian criticism of the bureaucratic 'rigidity' characteristic of industrial organizations (there comprised the ever-outdated character of normalized quality control testing procedures as a core chapter) so dear to Austrian and Chicago School radical liberal versions of neoclassical economic analysis.

10. The 'spiral' of regulation and innovation-driven market competition characteristic of the international industry for advanced financial services (Merton 1995a) is perhaps only paralleled by the hyper-complex reflexive dynamics described by the emergence of digital network security standards, under fire from hacker attacks. In the world of computer security research and development, the well-known ambivalent character of the computer 'hacker' seriously compromises the moral separation border between constructive forensic criticism and destructive informed forgery (Hollinger 1991). If the technical complexity of the standard repertory of forensic authentication trials developed by government security agencies for the surveillance of telephone and computer communications has improved dramatically during the past ten years, it has been largely as a learning by-product of *ad hoc* public prosecution actions conducted in the face of ever more sophisticated new forms of network-computer fraud and forgery (Shimomura and Markoff 1997).
11. For a sociological analysis of the strategic games of counterfeiting being traditionally played between the detective and the criminal (Kaye 1995), the scholar critic and the scholar rogue (Grafton 1990), and, more generally, between the public expert and the reflexive forger, see the excellent book by Bessy and Chateauraynaud (1995).

In ratifying their belief that, if uncertain as to its ultimate aggregate economic outcomes, a minimum of publicly induced procedural standardization in the market for proprietary financial risk control systems is always better than 'pure' market competition, freed from any type (direct or indirect) of macroeconomic controls, the reply of regulators is, on the other hand, patterned under the no less ancient grammar of engineer-type denouncements of market 'whims' understood as coordination 'failures' (Boltanski and Thévenot 1991: 334). In fact, far from accomplishing their intended objective, that is, reducing aggregate levels of risk in contemporary financial markets, the most probable effect of this new repertoire of 'normalized', 'clear', 'simple' and 'fast' meta-statistical tests of model risk as global financial supervisory tools would be the improvement of the *industrial quality* of applied financial econometric models.

In the 'exponentially innovative' environment of contemporary capital markets (Merton 1995b), it is indeed increasingly difficult to identify and distinguish, from an a priori, theoretical point of view, the kind of behaviour we would otherwise label as 'smart', 'reckless' or overtly 'criminal'. That is, successful research on the dynamics of innovative behaviour, like that on the management of high-risk technologies, depends on the exploration of a more fundamental theoretical topic: how to attribute merits and blames – or even legal responsibility and 'authorship' on an individual basis – in a social environment where 'chance' is always *a little too probable*.

It has been argued (Meier and Short 1983) that since modern industrial life is inherently risky and for this very reason the connection between purposive action and observable social consequences is radically ambiguous, it would always be controversial to point the finger at some particular type of social risk (financial risk) as the outcome of criminal conduct (financial fraud and forgery). And all the more so when, as is the case with the financial services global industry, innovativeness is central to responsible behaviour. My ultimate claim in this article is therefore that the inverse should also hold, that because modern empirical social science is also a high-risk enterprise, it should be deemed no less controversial to dissolve any suspicion of criminal conduct (financial fraud and forgery) into the ambiguous dustbin of unintended scientific error (model risk).[12]

12. In the words of a prominent expert in the field, efficient policy rules against scientific misconduct must 'be able to distinguish error from fraud, unintentional and even careless mistakes from intentional misconduct, and misstatements from deceptive misrepresentation' (Bernardine Healy, Director, National Institute of Health, 1991–93, Statement, 1 August 1991, at the *Hearings on Scientific Fraud* conducted by Congressman John D. Dingell's House Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, cited in Kevles 1998: 306).

Acknowledgements

This article draws on research materials previously presented at different fora. I gratefully thank Jens Beckert (Free University, Berlin), Richard Swedberg (Stockholm University), Michel Callon (CSI, Paris), Zoltan Szanto (University of Budapest), and Mirjam Sent and Arjo Klamer (EIPE, Rotterdam) for giving me the opportunity to present my research to different audiences. Initial drafts benefited from lively conversations with Antonio Escohotado and sharp criticism by Fabian Muniesa. This version has benefited from valuable comments by Phil Mirowski and Laurent Thévenot, and from the general discussion at the 'New Economic Sociology in Europe', 2000 Conference at Stockholm University. For all legible English expressions found in the text, the reader must thank Richard Swedberg, who kindly and patiently edited the final version. I am responsible for any remaining obscurities.

References

Abolafia, Mitchell (1996) *Making Markets*, Cambridge, MA: Harvard University Press.

Basle Committee (1988/1998) *International Convergence of Capital Measurement and Capital Standards*, Basle, July.

—— (1996a) *Amendment to the Capital Accord to Incorporate Market Risks*, Basle, January.

—— (1996b) *Supervisory Framework for the use of 'Backtesting' in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*, Basle, January.

Bernstein, Peter L. (1992) *Capital Ideas*, New York: Free Press.

Bessy, Christian and Chateauraynaud, Francis (1995) *Experts et faussaires*, Paris: Métailié.

Boltanski, Luc and Thévenot, Laurent (1991) *De la justification*, Paris: Gallimard.

Callon, Michel (1991) 'Réseaux technoéconomiques et irreversitibilités', in Robert Boyer, Bernard Chavanee and Olivier Godard (dirs), *Les figures de l'irreversibilité en économie*, Paris: Editions de l'EHESS, pp.195–230.

Chew, Lillian (1994) 'Shock treatment', *Risk*, September: 63–70.

Crane, Dwight, Froot, Mason, Kenneth, Scott, Perold, André, Merton, Robert C., Bodie, Zvi, Sirri, Eric and Tufano, Peter (1995) *The Global Financial System*, Boston, MA: Harvard Business School Press.

Derman, Emanuel (1996a) 'Model risk', *Risk*, May: 34–7.

—— (1996b) 'The value of models and modelling value', *Journal of Portfolio Management* 22: 106–44.

Dunbar, Nicholas (1998) 'The accord is dead – long live the Accord', *Risk*, October: 9.

—— (1999) 'This is the way the world ends', *Risk*, December: 26–32.

—— (2000) *Inventing Money*, New York: Wiley.

Falloon, William (1995) '2020 Visions', *Risk*, October: 20–2.

—— (1998) 'Rogue models and model cops', *Risk*, September: 24–31.

Gibson, Rajna, Lhabitant, Françoise-Serge, Pistre, Natalie and Talay, Denis (1998) 'Interest rate model risk: what are we talking about?', HEC Lausanne, WP no. 9803.

Giddens, Anthony (1993) *Consecuencias de la modernidad*, trans. A. Lizón, Madrid: Alianza.

—— (1995) *Modernidad e identidad del yo*, trans. J.L. Gil, Barcelona: Península.

Grafton, Anthony (1990) *Forgers and Critics*, Princeton, NJ: Princeton University Press.

Guldimann, Till (2000) 'The story of Riskmetrics', *Risk*, January: 56–8.

Hollinger, Richard (1991) 'Hackers: computer heroes or electronic highwaymen', *Computers & Society* 21: 6–17.

Irving, Richard (1996) 'Banks grasp VaR nettle', *Risk*, June: S16–S21.

Izquierdo, A. Javier (1999a) 'Techno-scientific culture and the Americanisation of international financial markets', paper presented at the 4th European Conference of Sociology, Amsterdam.

—— (1999b) 'De la fiabilidad', Ph.D. dissertation, Departmento de Cambio Social,Universidad Complutense de Madrid.

Jacobs, Bruce (1999) *Capital Ideas and Market Realities*, London: Blackwell.

Jorion, Philippe (1995) *Big Bets Gone Bad*, New York: Academic Press.

—— (1997a) *Value at Risk*, Chicago: Irwin.

—— (1997b) 'In defense of VaR', *Derivatives Strategy*, January.

Kaye, Brian (1995) *Science and the Detective*, New York: VCH.

Kevles, Daniel (1998) *The Baltimore Case*, New York: Norton.

Mandelbrot, Benoît (1997a) *Fractals and Scaling in Finance*, New York: Springer.

—— (1997b) *Fractales, hasard et finance*, Paris: Flammarion.

—— (1999) 'A multifractal walk down Wall Street', *Scientific American*, February: 70–3.

Meier, Robert and Short, James (1983) 'The consequences of white-collar crime', in H. Edelhertz (ed.), *White-Collar Crime: An Agenda for Research*, Lexington, MA: Lexington Books, pp. 23–49.

Merton, Robert C. (1995a) 'Financial innovation and the management and regulation of financial institutions', *Journal of Banking and Finance* 19: 461–81.

—— (1995b) 'A functional perspective of financial intermediation', *Financial Management* 24: 23–41.

Miller, Peter and Rose, Nicholas (1990) 'Governing economic life', *Economy and Society* 19: 1–31.

Millman, Gregory J. (1995) *The Vandals Crown*, New York: Free Press.

Mirowski, Philip (1990) 'Learning the meaning of a dollar: conservation principles and the social theory of value in economic theory', *Social Research* 57: 689–717.

—— (1991) 'Postmodernism and the social theory of value', *Journal of Postkeynesian Economy* 13: 565–82.

Porter, Theodore (1995) *Trust in Numbers*, Princeton, NJ: Princeton University Press.

Sent, Mirjam (1998) 'Engineering economic dynamics', in J. B. Davis (ed.), *New Economics and its History*, London: Duke University Press, pp. 41–62.

Shapiro, Susan P. (1987) 'The social control of impersonal trust', *American Journal of Sociology* 93: 623–58.
Shimomura, Tsutomu and Markoff, John (1997) *Takedown*, trans. H. Silva, Madrid: Aguilar.
Steinherr, Alfred (1998) *Derivatives: The Wild Beast of Finance*, New York: Wiley.
Stix, Gary (1998) 'A calculus of risk', *Scientific American*, May: 92–7.
Swary, Itzhak and Topf, Barry (1993) *La desregulación financiera global*, trans. Eduardo L. Suárez, México DF: FCE.
Taleb, Nassim (1997a) 'Interview: the world according to Nassim Taleb', *Derivatives Strategy*, January.
—— (1997b) 'Against value at risk', unpublished paper, electronic copy available at http://pw1.netcom.com/~ntaleb/jorion.htm

Javier Izquierdo is Associate Professor of Sociology at UNED, the Spanish National University for distance education in Madrid. He has published works on economic sociology, the sociology of economics and the history of stochastic concepts in the social sciences. He has a Ph.D. in sociology from Universidad Complutense de Madrid and wrote a dissertation on the rise of financial engineering and derivatives-driven financial catastrophes. A book based on this dissertation, *Delitos, faltas y Premios Nobel* (*Crimes, Misdemeanours and Nobel Prizes*) will appear in Spanish in 2001.

Address for correspondence: Departamento de Sociología I, Facultad de Ciencias Políticas y Sociología, Universidad Nacional de Educación a Distancia, Senda del Rey s/n, 28040 Madrid, Spain. E-mail: jizquier@poli.uned.es