



Research Paper

A Darwinian view on internal models

Paul Embrechts

RiskLab and Swiss Finance Institute, Department of Mathematics, ETH Zürich,
Raemistrasse 101, 8092 Zurich, Switzerland; email: embrechts@math.ethz.ch

(Received September 24, 2017; revised September 26, 2017; accepted September 26, 2017)

ABSTRACT

Current discussions on regulation within banking (Basel III and IV) as well as insurance (Solvency II and Swiss Solvency Test (SST)) are partly dominated by the way in which internal models may be used for the calculation of regulatory capital and solvency ratios. In this paper, I review these discussions from a historical, personal and academic point of view. I pay particular attention to the internal versus standard model debate, taking account of the various technological changes currently impacting on the financial industry worldwide.

Keywords: Basel Committee; internal models; model risk; quantitative risk management; robustness; standard models.

1 INTRODUCTION: THE EARLY DAYS

First of all, it is a great pleasure for me to be able to contribute to the twentieth-anniversary celebration of *The Journal of Risk*, which prompts me to start by detailing my own publication-footprint relating to the journal. No doubt the year 1998 was a particularly memorable one: the downfall of Long-Term Capital Management (LTCM) raised serious concerns about the financial industry's capabilities in controlling the risk embedded in complicated arbitrage models. I still vividly recall learning the details of LTCM's rescue while participating in a risk-related conference at the

University of Cambridge. Around the same time, Resnick, Samorodnitsky and I published a paper in *Risk* magazine (Embrechts *et al* 1998) stressing the need for industry members to start looking more seriously “beyond the bell-curve”, ie, considering non-Gaussian models that allow for a more realistic view of heavy tailedness to assess risk in financial markets. This makes them better suited from a quantitative risk management (QRM) point of view. As early as 1997, I published a book along with two co-authors, Klueppelberg and Mikosch, with the telling title *Modelling Extremal Events for Insurance and Finance* (Embrechts *et al* 1997), in which we provided the mathematical theory for extreme events – extreme value theory (EVT) – needed to look beyond the Gaussian horizon. On June 5, 1997, I gave a talk on the topic of *Modelling Extremal Events* at the Annual Risk Conference in Chicago. I still fondly remember that, at that meeting, I held the first physical copy of the book in my hands and proudly showed it to the audience.

Shortly after, Risk Waters Group asked me to edit a volume within their Risk Books series, fully devoted to the topic of extremes in integrated risk management. *Extremes and Integrated Risk Management* (Embrechts 2000) was to be the title. Browsing through its chapters today, I feel proud about the visionary content provided by many excellent contributors. In the late 1990s, Basel II was on everyone’s mind, especially the discussions around internal models. It was then that operational risk (OpRisk) was introduced as a new risk category. Several years later, two coauthors and I provided the first paper of the newly established *Journal of Operational Risk* precisely on the topic of (very) extreme events within OpRisk. It was entitled “Infinite mean models and the LDA for operational risk” (Neslehova *et al* 2006). Little did we know back then that, a decade later, operational risk losses would pale many other banking losses on Wall Street. From today’s point of view, the inclusion of OpRisk under Pillar I of Basel II may be seen as excellent foresightedness. Even then OpRisk was perceived as an intrinsically important risk category. One further reason for its inclusion was that regulators wanted compensation for the anticipated reduction in regulatory capital for market risk (MR) and especially credit risk (CR) at a global industry-wide level.

The above examples show the strong involvement that my colleagues and I had with the early publication outlets of the Risk Waters Group, later to become part of Incisive Media. It is not accidental that around the same time – on October 7, 1994 as a matter of fact – at the Department of Mathematics of ETH Zürich we founded, in collaboration with the Swiss financial industry, RiskLab (see www.risklab.ch). Through JP Morgan and RiskMetrics, value-at-risk (VaR) had just been introduced to the banking industry, partially in response to the famous 4:15 Weatherstone report. The then CEO of JP Morgan, Dennis Weatherstone, wanted (daily, by late afternoon – hence 4:15) an answer to the simple question: how much could JP Morgan lose if tomorrow turns out to be a bad day? In this context, at RiskLab, we discussed very early on the pros and cons of VaR-type risk measures. Topics such as the netting of

derivative positions and internal model approaches were being pushed by the financial industry, leading to interesting mathematical research problems. All this activity led to increased discussions between academics and practitioners, ie, to an increased link between academia and the professionals working in financial institutions.

Risk publications played a pivotal role in this interchange. At RiskLab, our aim was (and still is) to provide a discussion platform where relevant technical issues in the realm of QRM can be discussed openly. We coined the phrase: “RiskLab offers a **precompetitive** discussion platform for QRM-related problems.” Some of these discussions led to scientific publications that are now part of standard QRM technology. Beyond the introduction of EVT within QRM, it is worth mentioning two more relevant examples: my paper (coauthored with McNeil and Straumann) on models for dependence concepts between underlying random variables or risk factors going beyond linear correlation (ie, copulas (Embrechts *et al* 1999)) and the seminal paper on coherence of risk measures by Artzner *et al* (1997).

Extended versions of these papers were published in research journals, but practical summaries appeared early on in *Risk* magazine, very much facilitating the dissemination of the fundamental research-driven results throughout the industry. Both papers were also included in *Extremes and Integrated Risk Management* (Embrechts 2000), making up a section on “Risk measures and extreme value theory”. In these publications, one also learns about the potential pitfalls – and hence limitations – associated with VaR as a risk measure and the advantages of assessing risk by way of a coherent, sub-additive risk measure such as expected shortfall (ES). Hence the resulting push from within RiskLab for the adoption of ES as a regulatory risk measure instead of VaR. The debate about which risk measure (VaR or ES) to use in practice and for what specific purpose is still ongoing. The dialogue and interaction between professionals and academics has been further promoted and nurtured through the organization of a yearly Risk Day at the ETH RiskLab. Indeed, some of the aforementioned fundamental research and ideas were presented to the (local) financial industry at the kickoff event in 1998 (see Embrechts 1998). These early publications, along with further development and results accumulated in subsequent years, were summarized in standard QRM textbooks by McNeil *et al* (2005, 2015). The content of these books is supported and complemented through a webpage (www.qrmtutorial.org), where most R-program-based routines and algorithms used in the book can be found.

Next I will reminisce on the development of QRM ideas in finance and insurance, especially concerning internal models, from a current as well as a future perspective.

2 INTERNAL MODELS: A FIRST DIGRESSION

Ever since internal models came onto the regulatory scene, the world of banking and insurance has witnessed a “survival of the fittest” evolution of analytic models

across the various Pillar I product-type subcategories.¹ Within regulatory frameworks worldwide, eg, the international Basel accords (especially Basel II and Basel III) for banking and the European Solvency II and Swiss Solvency Test (SST) guidelines for insurance, larger (international) institutions were allowed – even encouraged – to come up with internal models that fitted their business profiles and product ranges in order to calculate regulatory capital. The institutions themselves very much lobbied for the use of internal models, citing their superior risk sensitivity. At the same time, they also aimed, quite naturally, for lower regulatory capital charges, eg, through diversification effects. The industry settled rather quickly on a fairly broad set of internal models to account for MR and CR, including models for standard products widely available and traded on the market.

Whenever a new product appeared, its introduction was followed, in a natural way, by further model developments and model selection processes. In the case of OpRisk, however, though many advanced models were introduced, this “evolution” or “race” to identifying the (most) appropriate or widely agreed-upon internal model under Basel II’s Pillar I did not really work well. There are many reasons for this: data scarcity, data quality and overall problem complexity, the extreme nonhomogeneity between the various risk subclasses of OpRisk, eg, ranging from internal fraud via external events to legal risk; and finally the Basel II guideline that Pillar I regulatory capital for OpRisk had to be calculated using a yearly 99.9% VaR. Insurance regulation (eg, SST) realized these fundamental issues early on, whereas banking regulation is only now getting to grips with them. Finally, the 2007–9 financial crisis provoked a regulatory onslaught against the use of internal models that is still ongoing. The discussions around the appropriateness, effectiveness, robustness and reliability of internal models in managing risk contributed significantly to a backlash against quantitative (eg, mathematical) models used within banking and insurance regulation. Adding to the lack of understanding and confusion (or perhaps trying to divert responsibility and distract from the sheer incompetence of the real culprits, intoxicated by a culture of excess, greed and irrational exuberance), a number of more popular publications wandered off to boulevard journalism rather than scientific seriousness, such as the Gaussian copula and how the underlying mathematical formula would have destroyed Wall Street (see Salmon 2009). At the time, even the

¹ Throughout the paper, I use Darwinian notions as metaphors, such as “survival-of-the-fittest model” or indeed the “model that best adapts to market change”. These played through my mind when, many years ago, I started reading the various Pillar I guidelines underlying banking and insurance regulation. Indeed, within the universe of internal models, industry was always looking for the best model. Having said that, I am surely not pushing for a deeper comparison, nor should this allusion to Darwinian concepts distract from the main themes of my discourse.

more serious financial press jumped on the let's-blame-the-quants bandwagon (see Jones 2009). The authors of these publications would have been well advised to have first read the thoughtful interview given by Steven Shreve, about one year earlier, on the topic of blaming (just) quants for the financial crisis (Shreve 2008):

The quants know better than anyone how their models can fail. For banks, the only way to avoid a repetition of the current crisis is to measure and control all their risks, including the risk that their models give incorrect results. On the other hand, the surest way to repeat this disaster is to trust the models blindly while taking large-scale advantage of situations where they seem to provide trading strategies that would yield results too good to be true. Because this bridge will be rebuilt, the way out of our present dilemma is not to blame the quants. We must instead hire good ones – and listen to them.

Shreve's statement is still highly relevant today, especially concerning the importance of quantitative (ie, internal) models. In fact, my own view in this respect was expressed in 2001, in an early response to the then new Basel II guidelines (Daníelsson *et al* 2001). In this paper, we very clearly put our academic finger on the weaknesses of the new regulatory proposals and very strongly warned against the possible disastrous consequences. Some of the weaknesses we highlighted included the neglect of endogeneity of risk, network vulnerability and systemic risk, possible procyclicality of the new guidelines, the data issue underlying OpRisk, the widespread (mis)use of rating agency AAA labels for complicated financial products, and some more technical suggestions on the use of risk measures (in particular supporting the use of ES instead of VaR). We summarized these concerns in the direct, crude style that is common to academics: "Reconsider before it is too late" (Daníelsson *et al* 2001, p. 5). (Unfortunately, but certainly not surprisingly, about five years later it was too late!) This academic response was officially submitted to the Basel Committee on Banking Supervision – a step I very much hope more of my academic colleagues will take. Whereas our 2001 paper received considerable attention and was widely read – and is still cited to this day – its immediate influence on the Basel II proposals was minimal. With hindsight, we could and should have done more. Our spot-on conclusions were based on findings gathered at a broadly attended conference on Basel II organized by the Financial Markets Group of the London School of Economics, and as such they were not merely based on ivory-tower thinking. Recognition for this and subsequent work came much later when Lord Adair Turner, then chairman of the UK Financial Services Authority (FSA), invited me to attend a conference in London on March 22, 2010, which was organized by the FSA. More precisely the invitation request read: "I would be delighted if you could join us at this event to speak about the modeling of traded assets and its role in prudential regulation." It is fair to say that some of the findings of our 2001 paper found their way, directly or indirectly, into future regulatory guidelines.

3 SOME RELEVANT QUESTIONS ...

In the current discussion concerning quantitative (internal) models in finance and insurance, here are some questions that need answering.

- (Q1) What precisely constitutes a model?
- (Q2) What is the role of calibration and how can models be validated, for instance, through statistical backtesting or, more precisely, through passing a series of appropriate controls based on criteria and tests that are part of a sound validation framework with respect to which models are evaluated and eventually accepted (or better formulated from a statistical hypothesis testing point of view, and not rejected)?
- (Q3) Where and how do model uncertainty and model robustness enter into the balance?
- (Q4) How do we communicate the ways in which models are used within an institution? How can the findings from these models be reported to the outside world, eg, in order to establish regulatory benchmarks and/or comply with Pillar II guidelines?
- (Q5) Respecting the anonymity of the data, how can the model itself, the fitting process and the backtesting be communicated to the academic community to gain insight into possible deficiencies early on? In a world where data is becoming more and more important, industry will have to become more willing to share data with academia. In some ways, industry has already learned a lesson by increasingly moving from licensed to open-source software, especially in the world of statistical data science.
- (Q6) How do we add stress testing components and, more importantly, which stress tests do we use?
- (Q7) What are the practical limitations and potential shortcomings of “one model to rule them all” and indeed of “one risk number (VaR or ES) to rule them all?” How do these possible concentrations impact upon systemic risk?
- (Q8) What are the conceptual and technical issues associated with going from a bank internal model landscape (or warehouse) to measures of solvency?
- (Q9) Is there a need for balancing or combining standardized approaches with approaches based on internal models?

- (Q10) What would be the consequences of the disallowance of the use of internal models for solvency purposes?
- (Q11) And (for the moment) finally: how should we view the internal versus standardized model dispute (and, more broadly, the need for more fundamental changes) in a business environment that is changing at an accelerated rate due to digitalization?

4 ... AND SOME ANSWERS

Concerning (Q10), it is clear that choices between standard and internal models mainly occur when it comes to the calculation and reporting of regulatory/solvency capital. Internal models are, and will remain, the rule in day-to-day risk management at the level of development, pricing and hedging of specific products. For capital purposes – and this by definition itself – standardized models typically allow for a better (ie, more congruent) comparison between financial institutions having a comparable risk or business profile. The word “comparison” is important here. More questionable is the lack of risk sensitivity to (especially adverse) market conditions. Concerning internal models used before and throughout the financial crisis, we have learned from regulatory studies across the wider banking industry that risk capital calculations for a given reference portfolio resulted in widely diverging capital numbers based on those models. As a consequence, risk capital numbers between banks became difficult to compare in an objective way. In an ideal world, the models used for setting capital should not be too different from those used for internal risk-management purposes. An overly marked dislocation between the two worlds of risk measurement cannot be our ultimate goal.

Concerning (Q11), the changes alluded to here are mainly driven by big-data management (better referred to as data science) and information technology (IT). Buzzwords include algorithmic trading, high-frequency finance, neural networks, machine learning, Blockchain technology, cryptocurrencies, distributed ledgers, smart contracts and the numerous developments in the FinTech/FinReg universe. In particular, the big-data developments herald a move from “know your client” to “know your data”, a move I personally am not very comfortable with. Below, I will comment only on some of the underlying issues. The proper and adequate assessment of the advantages and disadvantages of employing a standard or an internal model depends on various considerations (scope and purpose). For this and other reasons, there is no universal framework with respect to which we can identify a clear winner between standard and internal models (see also my comment under (Q10)). Their appropriateness depends crucially on the context and thus so does our choice

of model. Whatever solution one comes up with, a balance between risk sensitivity, simplicity and comparability must be the final goal.

One thing is for sure: in order for an internal model to become a gold standard in this changing and highly competitive technological environment, it has to be fully understood. There is no room for black-box magic. Models are tools designed to help us sharpen the questions being raised, potentially leading to increased understanding and support for decision makers. Thus far, models do not take decisions: people do. Perhaps we may now be witnessing exactly the transition from “thus far models do not” to “now models do”. For this reason, one has to be able to communicate the results from such internal models, as well as the underlying model assumptions, to a sufficiently wide audience in a clear, succinct and comprehensible way. Will companies – and not just models – seize the digital challenge and adapt? Clearly, the discussion about models and model development goes well beyond the realm of internal models, especially in view of the broad advancement of AI-based technology in all aspects of daily life.

In a personal communication with the author Enrique Loubet summarized the human-machine-decision process as follows: “It would be misleading and false to claim that models will ever be taking decisions. Most certainly, models will indeed become progressively more sophisticated as various steps in model development and model interfaces that currently require man-assessed choices and input (ie, steps involving human intervention and decisions) are likely to be automated and hence integrated into them. But this event will be a human decision in its own right. That is to say, we cannot hide our own responsibility in an automated process: to do so would be pretense. Humans, either as model developers, users of models or people taking decisions based on the output of models, are interacting with all aspects of the algorithms forming part of an ever increasing push for automation.” As it stands, however, the current discussion concerning regulation for insurance and banking is dwarfed by the much wider, more political and societal debate of what constitutes the ideal (or at least better) financial architecture of the future. The final lines of Jorge Luis Borges’s poem “Ajedrez” (or “Chess”) come to mind:

God moves the player, and he, the pieces.
Which god behind God starts the plot
of dust and time and dreams and agony?

In the context of the present discussion around internal models, we could conclude the following: internal models/quantitative models are embedded in corporate ones, and these in societal, ethical, political and governmental ones. That is the complex and intertwined environment in which we live, try to adapt and function. We may try to narrow the discussion to a component of this whole, only to realize that, when going deeper, we often can no longer remain in the isolated framework where we started.

5 A NEW ARCHITECTURE FOR FINANCIAL INSTITUTIONS AND ITS REGULATION

It is not surprising that politicians, academics and regulators, as well as the general public, want to rein in the complexity of banking institutions. See, for example, the discussions around limited purpose banking (Chamley *et al* 2012) and whether or not our bankers are indeed “wearing new clothes” (Admati and Hellwig 2014). At the same time, “classical” insurance and finance has become highly technical. This trend is expected to sharpen at an ever-accelerating rate. For insurance, for instance, it would surely be a bad move to turn away from market consistent valuation (MCV). Solvency I (so-called statutory) figures published by the Swiss regulator FINMA for life-insurance companies headquartered in Switzerland were hardly affected by the financial crisis at all – so much for risk sensitivity! Insurance regulation is very much based on policyholder protection and solvency guidelines aiming for an at-arm’s-length transfer of business between two parties in times of distress. At such moments in time, MCV is absolutely crucial, and properly worked out internal models (including results from well-chosen stress tests) offer key guidance here. This point of view is very much akin to the concept of a “living will” for banks. I used the term “classical” above deliberately; indeed, in a rapidly changing, increasingly technologically driven world, internal models, together with the intellectual capacity within companies and regulatory bodies, will prove to be critical. We all (regulators, industry professionals and academics) need to be able to attract the best minds in order to face up to current and future challenges. Whereas today the regulatory horizon is typically a year (though intermediate checks do of course take place), new products as well as the ambient IT landscape will move from periodic (regulatory) oversight to a much more dynamic form of supervision.

A statement to this effect was made by Nobuchika Mori, the commissioner of the Financial Services Agency of Japan (Mori 2016):

The safety and soundness of a bank cannot be captured by a point-in-time assessment of its balance sheet alone. They are ensured through dynamic interactions between the bank and the markets, and affected by various elements in the entire economy ... we intend to move from a framework dominated by static regulation to that complemented by dynamic supervision ... the global regulatory community aspired to maintain financial stability and enable sustainable growth by providing banks with **incentives to enhance their risk management practices**, capital strategies and business models. The JFSA is hoping to explore the potential benefits of such an approach once again.

The bold type for emphasis (which is mine) seems to hint at a possible rewarding of a proper use of internal models for risk-management purposes. Indeed, early on in his speech, Mori made the comment that:

the global regulatory community's preoccupation shifted from bettering risk management to enhancing capital adequacy. Less confidence is given to supervisory processes adapted to specific institutions and more hope is placed on the effectiveness of uniform rules. It is sometimes argued that the room for innovation in risk management can be abused by arbitrage and that regulators need to intervene deeper into banks' risk management processes.

The point on regulatory arbitrage is well taken, but it does not apply solely to the banking world. The reader is, for instance, encouraged to look at an example of shadow insurance, which grew out of regulatory optimization/arbitrage (see Foley 2016; Hepfer *et al* 2017; Kojien and Jogo 2016). In view of this and several other similar examples/constructions, it does seem strange that so much of current regulatory effort is aimed at moving away from internal-model thinking.

6 CAPITAL RATIOS AND REGULATORY ARBITRAGE

The calculation of regulatory capital constitutes the main battlefield between regulators and financial institutions when it comes to the use of either internal or standardized models. Regulatory capital always constitutes a quotient with capital in the numerator and a measure of risk in the balance sheet in the denominator. Hence, reporting higher regulatory capital numbers is not only achieved through “reducing” the risk-weighted assets, say, in the denominator, but also by “increasing” capital in the numerator. Both words – reducing and increasing – are in quotation marks as history has shown that a combination of clever financial engineering, creative accounting and tax optimization may be misused to this effect. Banks, as well as insurance companies, deliver products such as loans, client portfolios, risk hedges, alternative risk transfers, life- and non-life-insurance products and pension solutions; hence proper quality control of such products (as in the manufacturing industry) can be expected. Internal models become eminently important to assess and report the risk embedded in these products and to support the communication with clients of the potential benefits and risks associated with their transfers.

A well-balanced point of view on the wider debate was given by Isabelle Vaillant, the director of regulation at the European Banking Authority (EBA). Her summary statement is that the Authority's “key goal” in Basel talks has been to defend a risk-sensitive capital framework. In particular: “Non-risk sensitive items should be a non-dominant part of the framework.” And further: “It is not a mystery that the EBA has been defending **models** with the idea that **risk sensitivity is crucial**” (Wood 2017).

For the nonspecialist, it is definitely useful to follow up on some concrete discussions between industry and regulation on the topic of internal models. For example, in the context of market risk, one place to start is EBA's December 14, 2015 document

EBA/CP/2015/27, together with several industry responses, including those of the International Swaps and Derivatives Association and the Association for Financial Markets Europe. EBA's final draft, submitted to the European Commission, came about on November 22, 2016 as EBA/RTS/2016/07, a 127-page document. A somewhat broader scope on the use of internal models is found in EBA's February 28, 2017 document, a Guide for the Targeted Review of Internal Models (TRIM) (155 pages). The following statement from the foreword of the latter document is telling in light of the ongoing debate between industry and the regulators:

The Targeted Review of Internal Models (TRIM) is aimed at enhancing the credibility and confirming the adequacy and appropriateness of approved Pillar I internal models permitted for use by systemically important financial institutions (SIFIs) when calculating own funds requirements. The category of SIFI institutions was introduced in the wake of the financial crisis in order to better safeguard the global financial system against the threat of systemic risk. As a major objective, TRIM focuses on the **reduction of unwarranted variability in risk-weighted assets (RWA)** driven by inappropriate modeling which takes advantage of the freedom granted by the current regulation.

7 MODELS AND THE MARKET

Concerning models, it is difficult to beat the oft-quoted statement “all models are wrong, some are useful” by the statistician George E. P. Box. The full statement from his 1976 article “Science and statistics” (Box 1976), addressing the issue of parsimony, reads as follows:

Since all models are wrong the scientist cannot obtain a “correct” one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

His is a statement to be appreciated in the current deluge of machine learning and neural network technology. Whereas in the above quote one can – and, for the purpose of this discussion, one should – augment “natural” with “economic”, that broader interpretation has its consequences. First, most scientific models developed to describe natural phenomena are driven by incorporating sensible hypotheses from empirical observations. The scientific premises and predictions of these models can be tested and experiments repeated by other scientists to either corroborate or refute the results. Indeed, following Karl Popper, scientific models cannot be validated, they can only be falsified as soon as model predictions fail to match actual observations. In such cases, the models need to be reviewed and amended to extend their scope.

The financial market is a dynamic environment with multitudes of feedback and thus does not offer the possibility of repeating or running experiments under controlled conditions to assess models in the same peer-reviewed way as in science. For instance, in many ways we navigate in the dark when it comes to macroeconomics, a fact amply confirmed by several financial crises and the subsequent reactions of regulators, politicians and central bankers. How would one repeat or even test macroeconomic scenarios in an experimental economics environment? One particular difficulty stems from the fact that there exists a strong feedback to the market from models and the financial products that are being introduced; in a way, these define the “market” (see Ayache (2010) for a more philosophical discussion on this notion of “market”). Unlike in the case of most scientific models – except perhaps when dealing with small quantum scales – where the model and measuring devices do not distort the underlying physical phenomena in any significant way, financial models and active financial products directly impact upon and change the market. In other words, they define or affect the object they aim to describe. In finance, model development is driven by idealized assumptions on the functioning of a market – and on the agents participating and interacting within that market – often mathematically glued together with no-arbitrage arguments. It is not unrealistic to claim that, when it comes to the world of banking and finance, opacity reigns, with little data to guide us. Nonlinear effects appear in many, if not most, financial derivatives, and intricate dependencies compound to make precise modeling a very hard task indeed.

A capital adequacy ansatz – be it based on a standard or an internal model – always assumes that certain conditions pertain to the underlying markets. These conditions are (or should be) more obvious in the case of an internal model due to the explicit quantitative modeling. However, standardized approaches are also based on assumptions, at least involving the use of specific accounting rules. Browsing through the various standard models in Basel III, say, reveals a nontrivial amount of complexity and (often implicit) underlying assumptions. In both cases it is essential that assumptions are made explicit. They should be clearly communicated to the various stakeholders involved and adhered to throughout, up to the actions based on the conclusions they lead to. As an example of such discussions, consider the important case of OpRisk. After several years of an advanced measurement approach (AMA) ansatz for the modeling of OpRisk, the Basel Committee has recently decided against the use of internal models while supporting a standardized approach instead. In recent years it has become clear that great variability toward the calculation of OpRisk regulatory capital resulted from the different AMA models (and hence assumptions) used by larger international banks. The proposed standard model, however, also involves assumptions that need to be made explicit. For a detailed discussion on this, see Peters *et al* (2016), which includes references to the relevant Basel Committee documents.

8 REGULATORY ARBITRAGE, RATING AGENCIES AND PRODUCT COMPLEXITY

Whichever approach one takes, regulatory arbitrage is to be avoided and for this regulatory vigilance is needed. The unfortunate example of prime negative was a development of the collateralized debt obligation (CDO) and credit default swap (CDS) markets leading up to the 2007–9 financial crisis. All too often an AAA label was most gladly (and passively) accepted without questioning the rating agencies' (internal) models used to arrive at this label (at this point, reread the statement on quants made by Shreve earlier in the paper; see also the 2001 warning published by myself and my coauthors about these practices (Danielsson *et al* 2001)). Further, AAA ratings were typically interpreted as meaning risk-free, leading to massive buying and, more dangerously, bank-internal warehousing of such products. In doing so, banks as well as some insurance companies – typically with the suffix FP (standing for “financial products”) to their name – violated their life-long *raison d'être* of providing maturity transformation and the acquisition and selling of risk through properly diversified portfolios. As a consequence, by 2007, the financial industry as a whole was (akin to being) long an economic catastrophe bond waiting to be triggered by a macro event that, in the end, turned out to be a substantial decline in the American housing market (Acharya *et al* 2009; Coval *et al* 2009). The ensuing spillover of problems from Wall Street to Main Street is to this day present in our minds. Regulatory arbitrage between the banking book and the trading book allowed for a considerable reduction of internal-model-based risk capital charges.

A further issue that surfaced in the crisis concerned the calculation of RWAs, where minor model “corrections” led to considerable reductions in the risk capital (required to be) reported. An unsavory example of the latter is to be found in the case of the so-called London Whale. Read, for instance, the story as told in Jacque's excellent book, *Global Derivative Debacles* (Jacque 2010), especially the section on “The art of concealment” (pp. 300–4; see also my comments above related to EBA's TRIM document). One may add to this list Lehman Brothers's legal arbitrage of their pre-default leverage ratios using a REPO-105/108 accounting maneuver as well as the misselling of opaque financial products by several investment banks. All too often, the originators of such products were vague to say the least about the societal value of some of their products. Neither did they always understand them fully themselves. As a consequence, they were unable to communicate the underlying risks clearly to the eventual buyer.

After the crisis, this fact led to numerous fines being levied on the financial industry; fines that were then booked under the regulatory denominator of OpRisk, and this within the category of legal risk. By late 2006, the cumulative legal fines to the banking industry amounted to more than US\$220 billion. In terms of market capitalization of

individual (mainly investment) banks, values from 10% to 20% – even 50% in the case of Bank of America – were reported (see *Economist* 2016). It is therefore no wonder that observers started questioning the ethical standards and lack of social responsibility of bankers. The current “attack” on internal models is just one consequential aspect. Another major issue, especially for smaller countries with several large, systemic international banks and insurance companies, is that the regulatory inspection of the internal models used at such institutions is very costly, ie, putting an additional stress on capital and human resources. This is an observation that is difficult to bypass. It brings us back to the discussion on the increased sophistication of models and overcomplexity versus the actual relevance and societal value of many of the financial products being produced and sold.

No doubt quantitative models in financial risk management have brought considerable success to the economy as a whole and one cannot even start thinking of getting rid of such tools for trading, risk management and solvency purposes, especially at a time where innovative alternative risk-transfer solutions are increasingly in demand, eg, in the world of environmental risk. Internal models ought to be integrated into the core processes of financial and insurance institutions. This is one of the guiding principles underlying the Pillar II approach within Basel II/III. In the realm of insurance regulation, this corresponds to own risk and solvency assessment (ORSA (see, for example, Bernardino 2011)). ORSA, for instance, stresses the fact that there will always be a fine balance between (internal) models and leadership, but it draws the obvious conclusion that “models cannot replace leadership” (Bernardino 2011, p. 25). In regulatory practice – both for banking and insurance – forcing market participants to limit their choice of models must be avoided. Often, the Darwinian “best surviving model” does not exist. Thus, model diversity is not necessarily bad, as long as these alternatives (and their underlying assumptions and limits) are fully understood and communicated. The dangers of concentrating on one specific model in risk management were made abundantly clear during the 1987 crash, in part due to the widespread use of VaR-centered risk management and program-trading based on portfolio insurance (see, for example, Authers (2007) for a 2007 view on that crash). Recent losses in the realm of algorithmic and computer-program-triggered trading raise similar concerns. In times of crises, “cash is king” and liquidity risk surrounding models often comes unpleasantly to the forefront, causing everyone to panic and “run for the exit” (see Pedersen 2009).

9 RISK GOVERNANCE

Corporate governance ought to function in such a way that internal models are there to enhance the overall institution’s performance to the benefit of all stakeholders involved. It is nonsense to say (and here I quote an occasionally encountered criticism)

that “internal models do not capture tail risk” or that they cannot handle “complex interdependencies”; of course they can, but such model features must (and can) be included, often leading to higher capital charges. However, if the latter is the consequence, then so be it. One unpleasant consequence of the regulatory drive away from the use of internal models for capital and solvency calculations is that regulators may face difficulties in retaining their better quantitative people. Such a development cannot be healthy in the long run, whatever the financial architecture that will prevail in the future. Already, the worlds of banking and insurance are not easy ones to fully grasp. Add to this existing complexity ingredients such as interconnectedness (either social or technological), demographic changes and a major drive toward just-in-time production and delivery (with the emphasis on the pursuit of higher efficiency and productivity at the expense of increasing vulnerability – which has seen business interruption emerge as one of the main risk categories facing the world of insurance). There is also the whole drive toward a big-data-oriented society with almost continuous monitoring at all levels, and the ensuing threat of cybercrime and our increasing vulnerability to it – to name just a few of the changes just around the corner. With so many factors to consider, it becomes abundantly clear that we need “all hands on deck”. Any good risk-management department must contain a diversity of talent, skills, competencies and experience. Throwing out quantitative skills and reducing capital adequacy considerations to the predominantly qualitative/standardized level is necessarily incomplete and cannot be the ultimate goal. As Albert Einstein is often quoted to have said: “Everything should be made as simple as possible, but not simpler.” This is reminiscent of Occam’s razor or the Law of Parsimony we encountered in George Box’s earlier quote on models.

10 A SAMPLE OF SOME RESEARCH RESULTS

Independent of their immediate practical implications an immediate practical use, all the issues above (and many more besides) lead to fascinating research questions, questions that, ideally, should be tackled through a close collaboration and constant dialogue between industry and academia, including discussions with regulatory bodies worldwide. This brings us back to the spirit very much prevailing around the birth of RiskLab in 1994, as well as the early publications in *Risk* and the *The Journal of Risk*. In these early publications it was made abundantly clear that VaR, standard deviation and linear correlation are misleading risk metrics (to say the least) when it comes to applications in markets where the stochastic behavior of the underlying risk drivers is “well beyond the bell-curve”.

In 1999, McNeil, Straumann and I formulated, in a somewhat tongue-in-cheek manner, the First Fundamental Theorem of QRM (Embrechts *et al* 1999), stating that within the world of elliptical models (eg, the multivariate normal or multivariate

Student t) the three risk metrics above work fine; for a precise formulation, see, for instance, McNeil *et al* (2015, Theorem 8.28). At the time, however, we stressed the much more important Second Fundamental “Theorem” of QRM. In a nonelliptical world (ie, in reality), all the conclusions of the first theorem fail and, depending on how far reality deviates from the realm of elliptical models, the extent of the failure may be significant. In a nonelliptical world, VaR is noncoherent, standard deviation becomes a questionable measure of risk and linear correlation is not able to accurately capture dependencies. Unlike the first theorem, the second “theorem” is not one that is rigorously or precisely formulated. Nonetheless it summarizes numerous mathematical results from the world of model risk (Morini 2011) and robustness, which are fields of current research of considerable importance.

In view of some of these results, it is deplorable that VaR is still predominantly used for risk-management practice as the benchmark for managing financial risk (see Jorion 2006). VaR, as a high quantile of a profit-and-loss (P&L) distribution, is fine to report; actuaries and engineers have used this risk measure with great success for a very long time. In their language, this measure is referred to as a “once over a given time period return event”. Ample examples can be found in the reporting of earthquakes, floods and storm events. Statistical estimation of these risk measures remains difficult, especially at very high quantile levels and long return periods. However, engineers would never start adding up such return-level risk measures. Yet this is exactly what the typical practitioner in the financial industry does on a daily (if not a minute-by-minute) basis when VaR is applied to questions related to portfolio optimization, risk aggregation, diversification and allocation. Applied in such a way, it just becomes the wrong choice of risk metric in any realistic market environment.

If (and that is a big if) a risk metric is to be used as a P&L summary, then ES is far superior. Its general convexity properties (see Embrechts and Wang 2015) allow for its use in risk aggregation and allocation applications, to name just two examples. Equally important, by moving away from VaR- to ES-based risk reporting, one moves from an “if” to a “what if?”-oriented risk-management culture. The crucial point is that in answer to the question (raised in 2006, for instance) “What happens to our MBS portfolio if over the next two years American house prices fall by, say, 20%?”, an “if” reaction would be: the probability of such an event is astronomically small, it will not happen. On the other hand, in a “what if” discussion, one simply asks “What are the consequences for our MBS portfolio if that happens?” If the answer to that “what if” question is “We stand to lose several billion dollars”, then surely some managers or board members higher up in the hierarchy would (at least) raise their eyebrows. It was the enormous increase in volume and warehousing of perceived risk-free assets that created the financial crisis. The last example is based on actual facts; the real underlying example resulted in a US\$50 billion rescue plan. For more discussions

and examples as well as warnings on these issues, see McNeil *et al* (2015) or read some of the recent papers on my website: www.math.ethz.ch/~embrechts.

11 SUMMARY

Of course, I could have addressed many more aspects of the use and misuse of models in banking and insurance practice and regulation. Below I give a partial summary of the discussion so far, placing extra emphasis upon one or two points.

- (S1) I believe that internal models in the financial industry are here to stay, and that they need to be well understood and carefully documented as well as properly challenged and critically calibrated. Institutions are well advised to methodologically catalogue them and check for consistency of their usage across different product and reporting lines, resulting in what could be referred to as an internal-model book. At the level of calibration, the difference between internal and standard models may be blurred. A standard model ansatz may readily transform into a highly internal one at this stage.
- (S2) Internal models play an important role for regulatory purposes, as long as their use is methodologically clear as well as scientifically and ethically sound.
- (S3) I think it is important to understand changes in model-based capital values over certain time periods, instead of concentrating solely on their stand-alone values.
- (S4) I welcome the added value that results from combining standard risk-measurement procedures with internal-model-based ones; surely big discrepancies need reporting and require explanation.
- (S5) I strongly believe that modern IT may for the first time (after the proverbial “ATM technology” statement by Paul Volcker in 2009), push financial institutions in the direction of a very different business architecture; see for instance Shepherd-Barron (2017) for some historical comments related to “50 years ATM”. Indeed, the ATM was born on June 27, 1967 as Barclays Bank installed the first-cash dispensing machine (CDM) at its Enfield branch in North London. John Shepherd-Barron is broadly accredited as being the inventor of the CDM (see Shepherd-Barron 2017). As is often the case with new discoveries, related ideas for this technological development were prevalent at the time. In Philippon (2017) one can find a detailed analysis of why it took the financial industry such a long time to embrace the full power of modern IT, now referred to as FinTech. In Lo (2016), a blueprint is given on financial regulation in a FinTech environment. Finally, the associated legal challenges are broadly discussed in Arner *et al* (2017), leading to RegTech, which stands for technical solutions

to regulatory processes. As a consequence of all these IT-driven developments, the notion of an “internal model” may – and will – take on a whole different meaning in markets of the (not too distant) future, as surely as the already operational robo-advisor is different from our bowler-hat-wearing banker Mr Banks in *Mary Poppins*.

(S6) The determination of regulatory capital has important consequences for the emergence and growth of shadow markets as well as for the products (not being offered; the latter may even be more the case when this capital is solely based on standard models.

(S7) Corporate governance ought to function in such a way that internal models are there to enhance the overall industry’s performance to the benefit of all stakeholders involved.

As an appendix to the summary statements (S1) to (S7), I would like to quote from a very recent Risk.net publication (Osborn 2017), which arrived on my desk just before I mailed my paper to *The Journal of Operational Risk*. Its opening sentence sets the scene: “Banks are doubling down on the use of machine learning techniques for model validation in the face of regulatory skepticism over ‘black box’ methods.” One bank representative is quoted as follows: “Machine learning has proven particularly useful for validating the models built for regulatory stress tests, such as the Federal Reserve’s Comprehensive Capital Analysis and Review.”

However, from the regulatory side, one learns that:

Still, regulators remain wary about the use of machine learning in bank models. The Fed warned against using machine learning to assess contagion risk in model networks, saying these methods lack transparency and might obscure the true nature of banks’ vulnerabilities.

Be that as it may, the battle lines have been drawn and Darwinian evolution in the internal model universe will run its course.

As I already made clear at the beginning of the paper, I have thus far used Charles Darwin’s notions on natural selection and evolution in a somewhat light-hearted way. To conclude, however, in a more serious vein on the survival chances of good internal models, it seems appropriate to at least reflect more carefully on one of Darwin’s famous quotes: “It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is most adaptable to change.” One thing is for sure: there will be lots of changes (in modeling and beyond) that affect and perhaps reshape the worlds of banking and insurance. Industry, regulation and all the actors actively involved had better shape up for the change just around the corner.

12 EPILOGUE: 9/11

I started this paper with some personal historical notes. I would like to finish on a very personal one related to September 11, 2001. Earlier that year, I received an invitation to give a talk at the inaugural Waters Financial Technology Congress to be held on the morning of September 11, 2001, in the Windows on the World premises on floors 106 and 107 of the North Tower (Building One) of the World Trade Center in Manhattan, New York. Due to a clash with other commitments at the time, I reluctantly declined. I was already booked to give a talk on September 7 at the eleventh International AFIR Colloquium in Toronto on the topic of “Bounds on VaR for general functions of dependent risks: the actuarial approach” and needed to be back in Switzerland by that fatal day. At 8:46 on the morning of September 11, Flight 11 collided with the North Tower between floors 93 and 99. All sixty-five conference participants already present and sixteen staff members of the Risk Waters Group died; see Field (2002) for a very personal account of these events by the founder of *Risk* magazine, Peter Field. I would like to dedicate this paper to the memory of these victims.

DECLARATION OF INTEREST

The author reports no conflicts of interest. The author alone is responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

This paper grew out of discussions I have had on the topic of internal models over several years, with academics, practitioners and regulators. Early versions of my thoughts were presented at various international conferences. I take pleasure in thanking participants at these events. In particular, I would like to thank Enrique Loubet for his many thoughtful comments on an earlier version of the paper; his careful and critical reading very much convinced me of the importance of “critical thinkers” in our field. Finally, I acknowledge the excellent editorial editing of the paper.

REFERENCES

- Acharya, V. V., Cooley, T., Richardson, M., and Walter, I. (2009). Manufacturing tail risk: a perspective on the financial crisis of 2007–2009. *Foundations and Trends in Finance* 4(4), 247–325.
- Admati, A., and Hellwig, M. (2014). *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do About It*, updated edn. Princeton University Press.
- Arner, D. W., Barberis, J., and Buckley, R. P. (2017). FinTech, RegTech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law & Business* 37(3), 371–413.

- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997). Thinking coherently. *Risk* **10**(11), 68–71.
- Authers, J. (2007). The anatomy of a crash: what the market upheavals of 1987 say about today. *Financial Times*, October 18.
- Ayache, E. (2010). *The Blank Swan: The End of Probability*. Wiley.
- Bernardino, G. (2011). ORSA – the heart of Solvency II. Presentation, May 25, Groupe Consultatif Summer School.
- Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association* **71**(356), 791–99.
- Chamley, C., Kotlikoff, L. J., and Polemarchakis, H. (2012). Limited purpose banking – moving from “trust me” to “show me” banking. *American Economic Review* **102**(3), 1–10 (<http://doi.org/cd4h>).
- Coval, J. D., Jurek, J. W., and Stafford, E. (2009). Economic catastrophe bonds. *American Economic Review* **99**(3), 628–66.
- Daniélsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., and Shin, H. S. (2001). An academic response to Basel II. Special Paper 130, Financial Markets Group.
- Economist* (2016). Financial crime: the final bill. *The Economist*, August 11.
- Embrechts, P. (1998). Financial and insurance mathematics at ETH Zürich. Lecture, Risk Day at ETH RiskLab, September 25, ETH Zürich. URL: <http://bit.ly/2wP5Aux>.
- Embrechts, P. (ed) (2000). *Extremes and Integrated Risk Management*. Risk Books, London.
- Embrechts, P., and Wang, R. (2015). Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling* **3**(1), 126–140.
- Embrechts, P., Klueppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- Embrechts, P., Resnick, S., and Samorodnitsky, G. (1998). Living on the edge. *Risk* **11**(1), 96–100.
- Embrechts, P., McNeil, A. J., and Straumann, D. (1999). Correlation: pitfalls and alternatives. *Risk* **12**(5), 69–71.
- Field, P. (2002). Remembering September 11, the day I’ll never forget. Risk.net, September 2.
- Foley, R. J. (2016). Iowa at center of debate over “shadow insurance” deals. *San Diego Union-Tribune*, August 29.
- Hepfer, B. F., Wilde, J. H., and Wilson, R. J. (2017). Taking shadow insurance out of the shadows: regulatory arbitrage, taxes, and capital. Research Paper 2836215, Mays Business School.
- Jacque, L. L. (2010). *Global Derivative Debacles: From Theory to Malpractice*, 2nd edn. World Scientific, Singapore.
- Jones, S. (2009). The formula that felled Wall St. *Financial Times*, April 24.
- Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edn. Wiley.
- Koijen, R. S. J., and Jogo, M. (2016). Shadow insurance. *Econometrica* **84**(3), 1265–1287.
- Lo, A. W. (2016). Moore’s Law vs. Murphy’s Law in the financial system: who’s winning? Working Paper 564, Bank for International Settlements.

- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd edn. Princeton University Press.
- Mori, N. (2016). From static regulation to dynamic supervision. Keynote Address, April 13, 31st Annual General Meeting of the International Swap and Derivatives Association.
- Morini, M. (2011). *Understanding and Managing Model Risk: A Practical Guide for Quants, Traders and Validators*. Wiley.
- Neslehova, J., Embrechts, P., and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *The Journal of Operational Risk* 1(1), 3–25.
- Osborn, T. (2017). Banks tout machine learning amid regulatory concerns. Risk.net, September.
- Pedersen, L. H. (2009). When everyone runs for the exit. *International Journal of Central Banking* 5(4), 177–199.
- Peters, G. W., Shevchenko, P. V., Hassani, B., and Chapelle, A. (2016). Should the advanced measurement approach be replaced with the standardized measurement approach for operational risk? *The Journal of Operational Risk* 11(3), 1–49.
- Philippon, T. (2017). The FinTech opportunity. Working Paper 655, Bank for International Settlements.
- Salmon, F. (2009). Recipe for disaster: the formula that killed Wall Street. *Wired*, February 23.
- Shepherd-Barron, J. (2017). Meet the true star of financial innovation – the humble ATM. *Financial Times*, June 22.
- Shreve, S. (2008). Don't blame the quants. *Forbes*, October 8.
- Wood, D. (2017). Q&A: EBA's Vaillant on Basel IV, FRTB and CVA. Risk.net, April 4.

