July–September 2010 ISSN 1752-8887

Journal of

Risk Management in Financial Institutions

Volume Three *Number Four*

Contents

Henry Stewart Publications
Russell House, 28/30 Little Russell Street
London WC1A 2HN, UK

Henry Stewart Publications
Subscription Office, PO Box 361
Birmingham, AL 35201-0361, USA

www.henrystewart.com/jrmfi/

Journal of Risk Management in Financial Institutions Volume Three *Number Four* July–September 2010

July–September 2010 ISSN 1752-8887

Journal of
Risk Management in Financial Institutions

Volume Three *Number Four*

An official journal of PRMIA
The Professional Risk Managers' International Association

HENRY STEWART PUBLICATIONS

Validation of economic capital models: State of the practice, supervisory expectations and results from a bank study

Received (in revised form): 31st March, 2010

Michael Jacobs, Jr
is a Senior Financial Economist in the Credit Risk Analysis Division of the Department of International and Economic Affairs at the US Office of the Comptroller of the Currency.

US Office of the Comptroller of the Currency, One Independence Square, 250 E Street SW (Suite 3144), Washington, DC 20219-0001, USA
Tel: +1 917 324 2098; E-mail: michael.jacobs@occ.treas.gov

Abstract A challenge in economic capital modelling within financial institutions is developing a coherent approach to model validation. This has been motivated by rapid financial innovation, developments in supervisory standards (Pillar 2 of the Basel II framework) and the recent financial turmoil. Various practices are surveyed in validating economic capital models, both quantitative and qualitative approaches, and supervisory expectations and concerns regarding this process are discussed. The paper then illustrates several of these approaches (benchmarking, sensitivity analysis and testing for predictive accuracy) utilising data from major banking institutions' loss experience (from supervisory *call reports*), and estimates and compares alternative established frameworks for risk aggregation (including alternative copula models). Results suggest that practitioners may want to consider implementing a simple non-parametric methodology (empirical copula simulation (ECS)) in order to quantify integrated risk, in that it is found to be more conservative, as well as more stable than the other models, in a non-parametric bootstrap experiment.

Keywords: *risk aggregation, enterprise risk management, economic capital, credit risk, operational risk, market risk, copula*
JEL classification: *G10, G20, C10*

INTRODUCTION

Economic capital (EC) can be defined as a set of the methods or practices that financial institutions may use in order to consistently assess total risk in terms of a monetary measure known as capital, as well as attribute this measure among different risky economic activities. The origin of EC is as a means of allocating capital and assessing performance across different lines of business that might constitute a large, diversified financial institution. In that function, EC measures were adapted to perform *relative* risk assessments in a timely, reliable and accurate fashion. Understandably, there was little emphasis on measurement of overall risk as a capital figure.[1] Nevertheless, EC has evolved to a tool used in contexts in which accuracy in estimation of a cardinal risk measure is important, as is the case in the quantification of a bank's requirement with respect to the absolute level of internal capital in order to support its risk taking activities.[2] This development in how EC is deployed has been a function of both internal capital management needs of financial institutions as well as supervisory requirements and expectations. Furthermore, this has been facilitated by advances in EC modelling (ie theory and risk quantification methodology), and by new technologies (ie computing power and storage capability).[3,4]

While it is recognised that the field has largely come together over time in its notion of what EC is, such as a common understanding of key concepts across institutions having such models in place, it is noted that EC has since taken on a broader meaning. This can be seen happening at two levels: on the one hand, a reconceptualisation of the component risks to be aggregated into an overall EC framework, and on the other the relative prevalence and deployment of EC across the industry.

One may look at EC, how it is understood as well as utilised, at varied levels. This analysis of EC ranges from aggregation at the enterprise level, to the line of business or risk-type, and finally even beneath that to a portfolio of exposure. The components of EC are therefore complex, which presents a challenge for both practitioners as well as banking supervisors. To be more specific, there is Pillar 2 of the Basel II Framework, the process of supervisory review (Basel Committee on Banking Supervision (BCBS)).[5,6] This potentially involves a diagnosis by the bank and its supervisors of the *fitness for purpose* of the EC framework, known by the acronym ICAAP (*internal capital adequacy assessment process*).[2] If a bank has an EC model, part of the ICAAP involves a validation process with respect to that model. The contribution here is to survey and make recommendations of particular interest to supervisors and bankers where the validation of EC models is part of the supervisory dialogue. Furthermore, it is in the interest of supervisors to promote robust, transparent and effective risk management. This often requires an understanding of a bank's EC framework, which can be accomplished through a comprehensive validation programme. In any case, it is understood that ultimately EC is a tool of business, and as such designed and deployed by institutions for internal risk management purposes.

In this paper the importance of the role of model validation is emphasised on two levels. First, that validation promotes a grasp of how total EC is related to its components; and, secondly, that validation ensures that the components of risk are measured both consistently and coherently. The main part of this study is concerned with alternative means to validate overall EC processes, as opposed to the underlying risks measured by an EC model.

Among one of the more difficult areas in EC model validation is that related to frameworks for *risk aggregation* among financial institutions.[7–9] Methodologies in risk aggregation are considered less developed than the corresponding practices and techniques in risk measurement of the individual risk components.[10–12] Often such techniques are not coherent frameworks but rather *ad hoc* patchworks of shortcuts and judgmental overlays. Currently most banks either sum risks arithmetically, which implies the absence of *proportional diversification benefits*, or weigh such according to estimated correlations among risks, the latter being unrealistic in that it is unlikely that risks are generated by a multivariate distribution in the elliptical family. There are some cases where banks apply more elaborate methodologies, such as the method of copulas, or structural models that construct EC estimates from shared relations of risks to underlying a set of factors, but these are limited in scope.[2] Therefore, validation of EC models is problematic with respect to aggregation techniques employed in that framework. In particular, proportional diversification benefits — defined as the proportional reduction in diversified capital requirements vis-à-vis the case of perfect correlation[13] — implicit in models for inter-risk aggregation processes, such as estimates of dependency structures, are often based upon either expert judgment or industry benchmarks. Unfortunately, the accuracy of such factors has not or cannot be measured against the history or expectations of a bank, in that relevant data for such an exercise are generally lacking.

As EC models are often complex, having varied components, there may be little evidence that they are performing well or in line with prior expectations. An EC model may, moreover, incorporate assumptions regarding how variables are related or their individual dynamics, and as such may not hold in certain situations, such as stress or crisis periods. Validation is capable of giving internal and external users of EC output a level of confidence that model assumptions are reasonable. Furthermore, validation is of value in exposing the limitations of an EC model, in particular cases where underlying assumptions are not in line with economic reality.

EC model validation is currently in a young phase of development as compared to the validation of other types of models, such as methodologies for measuring market value-at-risk[14,15] or credit risk measurement.[12,16,17] Varied techniques exist with which to accomplish EC model validation, with each of these capable of informing as regards the robustness of only some of the desired model properties.[18,19] It should be noted that different validation techniques are powerful in the examination of only certain properties, whereby power is meant ability of a test to detect departures from a desired state of affairs. For example, many validation procedures exist that powerfully assess risk sensitivity (eg does the EC model reflect the state of the economy?), but as regards measuring overall absolute accuracy of EC (ie is an estimator of the loss distribution quantile predictively accurate?), few such techniques are available to do this reliably. Used in combination and particularly in an environment of robust controls and model governance, a broad set of validation techniques can offer more substantial evidence regarding performance of an EC model. Nevertheless, it is to be admitted that there still remains much scope for the industry to enhance the robustness of its validation practices that are informative regarding how well the EC models are calibrated overall, in particular for the cases where assessment of overall capital is an important application of the model (eg to set a financial institution's total risk 'budget'), as opposed to cases where it is not (eg where the model is used to establish the relative risks of different lines of business).

The second section of the paper will survey the different existing EC model validation practices in the banking industry, both qualitative and quantitative approaches, as well as discuss fitness for purpose of EC models and supervisory concerns regarding validation. The third section will illustrate a set of these quantitative techniques (benchmarking, sensitivity analysis and accuracy testing of an EC estimate) using the results of a bank study on risk aggregation. The final section will summarise major conclusions and discuss future directions for this research.

A SURVEY OF PRACTICES IN VALIDATION OF ECONOMIC CAPITAL MODELS

Validation can be interpreted in varied ways. In the narrow sense it is merely a statistical exercise, the *ex post* comparison of predictions to outcomes. On the other hand validation may be viewed more expansively as an evaluation of all aspects of a model, including aspects such as developmental evidence and analysis around the control environment. Herein, the latter broad interpretation is taken, where validation is a set of activities that encompasses all the processes that provide an evidence-based assessment regarding the *fitness-for-purpose* of an EC model.[20] This comprehensive assessment is likely to extend beyond aspects related to model performance, to things such as the quality of model management, or the systems environment within which an EC model is operated.[2] It is further recommended that validation processes are designed in conjunction with model development, as opposed to sequentially, while of course preserving the integrity of the model by maintaining the separation between developers and validators (eg having validators perform 'checkpoint' reviews at key milestones in model development).

The validation process can be thought of as having as its output evidence that a model is performing in line with expectations and standards. As EC models are often complex and multifaceted, incorporating several components that may not be in 'sync' at all times, it may be far from obvious in a reasonable time frame that the model is working well.[2] Such models incorporate assumptions (explicit or implicit) regarding how variables are related, or about the dynamics of such variables (eg statistical distributions of risk factors or correlations among a set of these). This can include the behaviour of economic variables during stress periods, which may not be incorporated into reference data sets, in which times such relationships may be expected to break down. EC model validation can give internal or external users (eg bank risk managers or banking supervisors, respectively) a measure of confidence that underlying assumptions have good properties (ie that they are reasonable, empirically grounded and conservative). It should be also noted that validation, as is broadly conceptualised, is a valuable aid in identifying limitations of an EC model, because even if fully validated no model is ever a complete depiction of economic reality.

As mentioned in the introduction, while certain aspects of the model validation present one with powerful tools (eg evaluating the sensitivity of EC to some measure of risk in the economy), it has less power for other aspects (eg measuring the accuracy of an estimator of high quantiles in a loss distribution). Nevertheless, the objective of the modelling exercise need not be restricted to predictive accuracy, as some EC models are developed as either analytical or decision support tools as opposed to historical data calibration (eg certain macroeconomic models).

It is argued that this holistic interpretation of validation is fully consistent with and grounded in the framework of the advanced internal rating based approach to measuring regulatory capital ('Basel II' or 'A-IRB').[6,21] It is true that this framework was couched in the context of developing the A-IRB parameters probability of default (PD), loss-given-default (LGD) and exposure-at-default (EAD), where assessment of the accuracy and reliability of these risk parameters is necessary to support calculation of minimum regulatory capital. It may be contended that validation of EC models is fundamentally different than that of IRB models, as EC output is a full loss distribution, in contrast to a predicted rate of default or of loss severity, which may be more amenable to backtesting analysis. The author thinks that this is partially true, but note that one may think in terms of a distribution of default or loss rates if one admits parameter uncertainty, as in a Bayesian framework.[22] On the other side, while estimating a loss distribution, the object of EC is the estimation of a moment of such a distribution, in particular a high quantile, so that analogies to the A-IRB model development may not be so strained (eg PD estimation in a low default portfolio setting[23]). While it is noted that conceptually EC models are most similar to value-at-risk ('VaR') models,[24,25] long used in the measurement of market or credit risk, there are several differences worthy of mention, such as the long time horizon (one year), high confidence levels (99.9th or 99.97th percentiles), and the scarcity of data (ie 'Have you or are you capable of ever observing enough three in 10,000 events for a reasonably-sized statistical sample?'). Consequently, validation methods used in practice can reasonably be expected to differ from those used for market or credit VaR.

Enterprise-wide internal EC models are not used as part of the Pillar 1 minimum capital requirement calculation, as there is a regulatory formula prescribed to banks.[5] Therefore, the concept of fitness-for-purpose takes prominence in the discussion around validation of EC models. EC is intended to be deployed variously depending on intended purposes that are likely internal to the institution, and validation seeks to assess whether such a model is adequate towards its intended purpose. It is also to be mentioned that the objectives of EC and regulatory capital differ, so it may be reasonable to expect that some details of their implementation may diverge.

Principle 1 of the BCBS validation principles states that 'Validation is fundamentally about assessing the predictive ability of a bank's risk estimates and the use of ratings in credit processes',[21] which is an explicit reference to assessing the predictive ability of a credit rating system, with an emphasis on the performance of a model as judged by the quality of its forecasts. Therefore, the state of affairs is that Principle 1 is a statement about rating systems. But a natural evolution of this principle in the context of EC modelling is that validation is concerned with the predictive properties of those models as well. Therefore, EC models are expected to incorporate forward-looking risk estimates, and as a consequence their validation is inseparable from an assessment of the quality of the EC estimates. It can be argued that such a restated principle remains appropriate in this context, as the validation processes and techniques presented and discussed herein all hold the promise of information regarding an expectation of the model's predictive ability, where this quality is broadly interpreted.

There are various other Basel II validation principles that are worthy of mention in this EC context.[21] First, the supervisory expectation is that a bank has primary responsibility for validation (ie outsourcing this to a consultant is not acceptable). Secondly, validation is an iterative process, occurring ideally in tandem with model development, not a 'one-off' exercise. Thirdly, there is no single technique or method that constitutes adequate validation, so that providing a range of tests is the standard. Further, validations should address processes of both quantitative and qualitative natures. Finally, validation should be subject to an independent review, ideally by a party that is both structurally and functionally independent of the unit responsible for developing and validating the EC model. Nevertheless, it should be noted that there is currently much debate in the bank-supervisory dialogue regarding the extent to which structural independence is an absolute necessity, driven mainly by resource constraints at banks. While it is believed that the notion of validation as expressed in this paper is consistent with those principles, an opinion is not put forward regarding which entity should be charged with validation, or who has the responsibility to sign-off on the end product.

This paper is mostly concerned with listing and describing the types of EC model validation processes either already in use, or potentially usable, by banks. It neither makes any warranty that such a listing is comprehensive, nor does it make a recommendation that banks use all or even some of these techniques. It only strives to illustrate the wide diversity of techniques that fall under the broad view of validation with respect to EC models, giving rise to what may be termed a 'layered approach' (ie the more layers that validators can produce, the greater the level of comfort that the validation can provide regarding mode performance). Furthermore, it sets out to show that each validation technique is informative with respect to only certain of the favoured model properties. In the subsections below, it is proceeded from the most to the least of processes characterised as qualitative, and then similarly for the quantitative counterparts. Along the way an opinion is expressed upon the character and extent of use.

Qualitative validation of economic capital models

The first qualitative validation process for EC that is discussed is the *use test*. While no longer explicit in the language of the US Basel II Final Rule,[26] the philosophy of the use test is fully incorporated and very much at the heart of the Basel II framework. In discussing the use test in the context of IRB, BCBS[21] stresses that, as part of the quality check process of IRB components, the use test is a necessary supplement to the overall validation process, as it has a critical role in ensuring and encouraging the accuracy, robustness and timeliness of such components. This is seen as a confirmation of the bank's trust in those components, which allows supervisors to place more reliance on their robustness and thus on the adequacy of regulatory capital. In the context of EC models, the use test is easily understandable, the notion being that if an institution is actually deploying risk measurement systems internally (and it may be added where there is impact upon profitability), then supervisors are likely to have more confidence in the model's output for purposes of computing required regulatory capital. A caveat here is that successful application of the use test requires a deep understanding of which aspects of a model are being used and how. For example, is the capital attributed by the model to a line of business being used in a transfer pricing mechanism, or is the relative riskiness of such units as assessed by the model an input into the bank's strategic decision process (eg acquisitions or dispositions of business lines). It can be judged that this is the most qualitative of the validation processes discussed here, as there is no assessment of any technical aspects of the EC model under consideration, but rather a pure evaluation of the 'business case' for the model.

It is common for banks to have in place a manner of *qualitative assessment process* with respect to their EC models,[2,4,18] as with other risk management models. This process could entail review of developmental evidence (eg model documentation, development work, formulae derivation or code), dialogue with key players (eg model developers, managers and users), comparison with external benchmarks (eg practices of comparable banks, industry studies and other publicly available information). This form of qualitative review is best able to answer questions regarding the conceptual underpinnings of the EC model. This includes the question of whether the model works in theory and, even beyond that, whether such theory is conceptually well-founded. Other examples of questions that qualitative assessment potentially addresses include the incorporation of economically correct risk drivers and if the mathematics of the model are right.

An EC model can also be qualitatively validated through analysis of *systems implementation*, and indeed this is rather widespread among banks.[4,18] It is a standard industry practice to subject any production-level risk-measurement system to rigorous IT testing prior to implementation. Such testing includes user acceptance testing, checking of model code and assessment of the IT control environment. The reason why these processes are potentially considered a component of the overall validation programme is that they would assist in an evaluation of the integrity of EC model implementation.

Management oversight is meant as referring to the involvement of senior management in the validation process.[2] This includes elements such as review of EC model output and the use of the resulting risk measures in making decisions at the line of business level (eg credit approval and pricing and setting of trade limits). Senior management are expected to be informed fully in model use, interpretation of model output and any limitations of the model. This should fully incorporate the specific implementation of the EC model.

Another form of qualitative validation in an EC context, not traditionally viewed by the industry as a form of validation, is *data quality checking*. Increasingly, these activities are moving to the forefront of supervisory thought in regard to validation of EC models.[4] These procedures refer to the activities that are meant to provide comfort with the completeness, accuracy and appropriateness of data used to develop, validate and operate the EC model. Qualitative validation processes falling under this category potentially include review of sourcing (eg data collection and storage), data cleansing processes (eg identification of errors), reviews of the extent of proxy data (eg use of equity market indices in lieu of trading revenue history to quantify market risk), review of any processes that need to be followed to convert raw data into suitable model inputs (eg scaling processes) and verification of transaction data such as exposure levels. Such a list is often a helpful indication of the level of understanding of the model.

Finally, for qualitative techniques in EC model validation, there is discussion on the *examination of assumptions*, which is also called *qualitative sensitivity testing*.[2,4] EC models, as with all models, rest upon various types of assumptions. Furthermore, some of these are hard-wired and cannot be changed without fundamentally changing the model. Certain of these assumptions are immediately evident, but some are far from obvious, while others are not so critical. For example, assumptions about certain fixed parameters (eg probabilities of default or recovery rates in credit; window lengths or proxy indices in market risk; or correlation parameters in risk aggregation or credit risk), distributions (eg for risk type, including assumptions about the shape of the tail distributions), or assumptions regarding the key economic agents (eg senior management, counterparty or customers' reactions to a crisis event) could be slightly varied. Banks vary in the degree to which they document and make explicit assumptions, including analysis of how the EC risk measure responds to changing these, as well as an assessment of any limitations that such assumptions place on how the EC model is used and applied. An example of the latter might be the implications of a choice for a *copula* to aggregate risks on the joint behaviour of risk types in the tail,[27–29] such as the well-known lack of positive or negative tail dependence in the commonly employed Gaussian copula. As will be seen in the next section, there is a similarity between the sensitivity testing discussed here and certain quantitative approaches; indeed, this aspect of examination of assumptions regarding inputs is the least (most) qualitative (quantitative) of the techniques described in this (the next) section. Furthermore, examination of assumptions could be included in activities described previously in this section on qualitative validation of EC models (eg use test, documentation of developmental evidence or an assessment of conceptual soundness).

Quantitative validation of economic capital models

The first type of quantitative validation is termed the *examination of input*s in the industry, and herein also as *quantitative sensitivity testing*.[2] As some parameters of an EC model are statistically estimated, the quality of such estimates — including sampling error — has a direct bearing on the accuracy of the risk output. Examples of econometrically derived inputs include the main IRB parameters PD, EAD and LGD in the credit component; correlations in the structural credit risk model or market VaR model component; and parameters of marginal distributions and dependency structures (copulae) in the operational risk components or in the risk aggregation framework for the EC model. It is understood that a full EC model validation should involve complete validation of these inputs, which in itself can be a broad validation as defined in this paper, including qualitative aspects such as things like use test, data quality, etc (ie it is a microcosm of the validation of the EC model). Quantitative techniques could include replication of estimates against historical data, benchmarking to alternative estimates (eg market implied quantities, rating agencies estimates or vendor models), and examination of outcomes-based validation (backtesting or out-of-sample analysis). Another part of this is assessing the materiality to EC model output through the sensitivity testing of parameters,[30] which would be a complement to the examination of assumptions and sensitivity testing described at the end of the preceding section on qualitative validation. Nevertheless, in comparing this activity to that qualitative form of sensitivity analysis, a difference lies in that, under the quantitative version, a structured approach to shocking the inputs is used (eg bootstrapping, resampling or simulation-based techniques). This produces a large range of EC outputs (or a distribution of the quantile estimator) with which probabilistic statements can be made.[13,31] For example, in assessing correlation inputs, one could shock them to some benchmark value (eg the Basel prescription, or a stressed 100 per cent) and look at the change in EC output, or one could look at a wide range of EC outputs by drawing from a distribution of correlations. This could be achieved through bootstrapping, if model parameters were re-estimated, or through Bayesian techniques, if a prior distribution for those parameters were formulated, and then simulated from that in order to obtain the posterior distribution of the EC estimator.[22]

It should be emphasised that this quantitative evaluation of EC model inputs, either looking at the estimation of inputs or measuring the impact of the later upon the output, is probably never going to be sufficient to achieve a full quantitative EC model validation. The problem is that the more complex or multifaceted the EC model is, then the greater the risk that there is a model error that transcends the quality of the inputs. If fundamental economic assumptions underpinning the model are wrong (eg the common assumption of no credit risk in structured market risk instruments, or of no liquidity risk in complex credit products, which were clearly big mistakes), then no matter how accurately estimated the parameters of the model are, or how insensitive the EC risk output to error in the inputs, there is no assurance that the model is validated. While the checking of EC model parameters is unlikely to illuminate in that regard, it is still possible, to some degree, to assess the accuracy and fitness-for-purpose of the model by deploying these techniques.

Model replication is defined as the exercise of trying to reproduce, as closely as possible, the risk measure of a bank's EC model. This involves a validator attempting to build a model of the same type, using the same reference data and applying them to the bank's portfolio. While this is considered a useful means of quantitative validation of an EC model, in spite of its expense, it is not widely used.[2] In its ideal implementation, the replication of EC model output would be a completely independent process, utilising independently developed algorithms and acquired data, which should mirror that used by the bank. Nevertheless, it is observed that, for practical purposes, EC model replication typically utilises some of the existing processes. Such leveraging of existing resources could be accomplished by running the bank's algorithms on an independently sourced data set, or alternatively utilising the bank's data sets in conjunction with independently derived algorithms. A caveat here is that, in either of these partial approaches, either the appropriateness of the banks' reference data, or accuracy and reliability of its algorithms, should have been first validated in order that conclusions of this replication might be considered meaningful. The value of model replication is that it gives rise to questions that challenge the extent to which developers or validators fully comprehend the data definitions, analytics and algorithms that underlie the EC model. For example, EC model replication can be used to identify errors in computer code, or to check that data queries are performed accurately. Nevertheless, as with all the techniques discussed in this paper, there is no expectation that the application of this technique constitutes full EC model validation, and as noted previously there is scant evidence of this being used widely by banks. Finally, it should be pointed out that a third party re-running the same algorithms, with the same reference data and on the bank's portfolio, is not considered true EC model replication — some of these elements should be independently generated (ie the replicator either rebuilds an EC model of the same type or attempts to generate the same type of reference data).

Benchmarking, also known as *hypothetical portfolio testing*, refers to either analysis of the comparability of a bank's (or banks') EC model output on either a given external benchmark portfolio; or alternatively, examination of a benchmark model on a bank's (or banks') reference portfolios.[18] In the former case, one has an internal risk management tool for *model risk management*, looking across EC models with respect to a given reference data set, thereby identifying outlier EC models (however, in this exercise parameters must first be harmonised for this to be meaningful). In the latter case, one has a supervisory tool of *bank portfolio management*, looking across banks with respect to a given reference model and providing a means of identifying outliers with respect to risk. In the context of credit risk capital, an example of the model risk perspective for a bank is comparing the results of an in-house model to either a vendor model (eg Moody's KMV Portfolio Manager™), the regulatory capital formula, or an academic model of credit risk;[32] in any case, after parameters have been standardised. An example from a broader context is the comparison of an internal structural (bottoms-up) model that integrates credit and market risk, and perhaps brings in another risk (like operational) using a proprietary copula model, with a benchmark such as a Gaussian copula applied to stand-alone credit, market and operational risks. In one example of a supervisory application of hypothetical portfolio analysis, one could have a set of banks required to run a composite portfolio of several banks (perhaps provided by a vendor or compiled by the supervisor) through their economic (or just credit, market or operational risk) capital models.

This validation methodology is considered powerful, in that it is readily capable of identifying outlier models or institutions, and is informative of attractive EC model properties such as relative risk quantification (or rank-ordering). Nevertheless, several limitations must be noted, chief among which is that benchmarking is only capable of relative comparisons among EC models. Unfortunately, this property provides cold comfort that any one model is most in line with reality, nor can it make statements regarding the accuracy of the absolute EC risk measure produced by any single model. It is recognised that it should be no surprise why outliers are seen in benchmarking or hypothetical portfolio analysis, as different EC models may be designed to work well in alternative situations (eg EC models that produce risk estimates that are stable through-the-cycle versus models that are designed to produce more point-in-time estimates). Furthermore, it is reasonable to expect that interpretation of this analysis would be complicated by differing economic foundations across EC modelling frameworks. Nevertheless, benchmarking is currently the most prevalent form of quantitative validation for EC models,[2] as there exist several points of comparison. These include industry survey results, vendor models (eg rating agencies), industry consortia, consultancies, academic studies and regulatory capital models. It is concluded that, while it may be challenging to gauge the level of comfort that may be derived from this activity, this technique still has value in confirming that either EC model inputs or outputs are comparable on a broad level.

Backtesting is now considered as a quantitative validation technique for EC models (or *outcomes-based analysis*), long considered the 'gold standard' of statistical validation in contexts in which data are sample (eg retail scoring models). It answers the question regarding how well a model's forecasts line up with realisations. This type of analysis takes several forms, such as tests of rank-ordering (eg area under the receiver operating curve), or tests of predictive accuracy (eg Hoshmer-Lemeshow chi-squared), and there is extensive literature on this topic. It must be noted here that recently the BCBS,[17] in the context of credit risk quantification, discusses the issue of *low power*. This means that, in cases of few observations of the outcome of interest, statistical tests may not be very informative about the performance of the model. A classic example is a low default portfolio setting (eg investment grade credits or commercial real estate), in which observed default rates may be zero or very small while PD estimates are positive (and in some cases rather large), so that the model is severely over-predicting (albeit it is 'conservative'); but the confidence bounds around observed default rates may be very wide owing to a small count (or undefined if zero), so one has no idea if the model is working. Said differently, under low power one is likely to be unable to reject any model. The case of EC is analogous to the extreme where no defaults are seen — for most portfolios, one probably may never have observed the one in 10,000 year event for the 99.99th percentile of loss (or if one believes that it was the recent crisis, one has at best one observation). Sadly, while there is some hope in the low-default portfolio credit context that there may be enough data in a few years, for EC it is probably the case that within our lifetimes there will never be sufficient to perform a meaningful validation.

Some researchers have suggested certain versions of backtesting which have improved power, which unfortunately are all subject to severe criticism.[2,4,20] One example is to perform more frequent validations, over shorter holding periods, such as a quarter or a month versus a year, or the use of overlapping holding periods (eg it is possible that a tail event occurs intra-year). Of course, that presents challenges in extrapolating these results to a one-year, non-overlapping holding period, the supervisory requirement. Another suggestion concerns the use of cross-sectional data, which involves backtesting across a range of reference portfolios, in that the aggregation of portfolios may result in losses cancelling each other out (ie one may get some or more tail events within sub-portfolios). The problem here is that then the ability to incorporate potential diversification benefits into the analysis is lost. Another proposal involves trying to extract information from a greater range of the loss distribution by backtesting at quantiles short of the 99.9th or 99.97th (eg 80th, 90th, 95th etc), which may provide enough data to say *something* about the adequacy of absolute risk measurement (eg if at the 90th quantile the model is deficient 20 per cent of the time, then it is evident that something is seriously wrong). But it may be argued that this does not add to the knowledge of the high quantile that is the object of interest. Finally, the most diehard critics will point out that backtesting is of use mainly where the model produces outputs that are quantifiable metrics (eg a default) for which an outcome can be potentially compared, and that an event which may never be observed within a reasonable time frame for which to perform this analysis does not satisfy this requirement.

The final quantitative technique to be discussed is *stress testing*. This is defined as analysis of and comparison to an extreme loss metric that in some way transcends the EC model design, which can be performed in two ways. First, model assumptions can be stressed (beyond what would be expected in 'normal circumstances') and then the EC model output examined, which — although by using the model in question — results in a loss that is outside the realm of the modelling framework in place for measuring EC. Alternatively, an estimate of stressed loss that is completely outside the model can be observed and compared to the EC produced by the model that is being validated. Such validation is capable of revealing weaknesses of the EC model and can identify circumstances in which one may face constraints upon EC. This can be thought of as a complement to the other validation techniques that have been discussed, a means of quantifying potential losses under circumstances beyond the range of history upon which the EC model has been constructed.

Comparing stress losses — either from a stress of EC model inputs or from an external estimate — allows a degree of confidence in the model-produced absolute risk measure. While the current stressing of Pillar one models is standard in the industry, there is little evidence at the moment of this practice with respect to EC models.[2]

Supervisory concerns regarding EC model validation

This section on surveying validation practices is concluded with a few supervisory concerns. Compared to just over a decade ago, a heightened supervisory focus on the validation of EC models can now be seen. As typified in the recent studies by the International Association of Credit Portfolio Managers (IACPM) and the International Securities and Dealers Association (ISDA),[19] it is EC model parameter harmonisation and cross-bank benchmarking which are viewed by regulators as aspects of model validation that have room for improvement. On the positive side, supervisors have the impression that banks are doing well with respect to quantifying risk on a relative basis (ie between lines of business or risk types) and producing models that exhibit sensitivity to risk factors (eg to the state of the economy). Supervisors recognise that, while there is quite a way to go still until these aspects are considered robust, banks have made progress in these areas, which is a source of optimism. Nevertheless, there are other areas in which the state of EC model validation is considered weak to the point where there is little hope. These are applications in which the model is used to assess overall EC adequacy of a financial institution, or where calibration to the bank's experience is important. As discussed in this paper, there is a common understanding in the industry that the validation of these aspects is fundamentally problematic and technically difficult, as in this context one is facing the quantification of high quantiles of a loss distribution, given limited data and over a long horizon. While supervisors accept that in some cases the EC model will be used in such a way that this is not a fatal flaw, it is understood that, when the focus of EC model validation is the assessment of capital adequacy, banks may find themselves in the situation of measuring and managing EC with poorly calibrated models. Therefore, the supervisory expectation is that bank model managers should be explicit about areas in which validation of the EC model is likely to be weak, and inform senior management that a validation cannot be fully executed. This kind of dialogue is necessary so that key stakeholders in this process clearly understand the heightened degree of uncertainty around EC model output in this situation, and are in a position to apply the appropriate degree of conservatism. As part of this, the potential losses associated with using a potentially misspecified EC model should be fully explored.

ECONOMIC CAPITAL MODEL VALIDATION EXAMPLE: ALTERNATIVE RISK AGGREGATION MODELS

In this section are illustrated several quantitative methods of EC model validation through presenting results from Inanoglu and Jacobs.[13] In that study the authors develop proxies for five risk types (credit, market, operational, trading and interest income) from historical quarterly call report data for five of the largest banks as of Q4 2008. Then they proceed to compare the EC output of different copula models for combining these according to absolute levels and variability. They use a non-parametric bootstrap to assess the accuracy of output estimation error in inputs, the parameters of marginal distributions (ie central tendency and the dispersion measures) and the dependency measures (ie correlations). While not a study of EC model validation *per se*, this illustrates several quantitative techniques discussed herein, including benchmarking/hypothetical portfolio analysis of alternative models, sensitivity analysis of EC model output to inputs and the testing of EC model accuracy through developing confidence bounds for high quantile estimates of the aggregate loss distribution. The main conclusion is that a non-parametric model for risk aggregation (the *empirical copula*)[33] is not only more conservative than commonly employed copula models (eg variance-covariance approximation or Gaussian), but is also less variable in the resampling experiment (more stable or accurate).

Table 1 summarises characteristics of the data set as of the fourth quarter of 2008 for the 200 largest banks (the 'top 200') in aggregate that represent a hypothetical 'super-bank' ('AT200') and individually for the top five banks in BVA (or the 'top five'). The five largest banks by BVA as of Q4 2008, in descending order, are as follows: JP Morgan Chase (JPMC) — BVA = US\$1.85tn, Bank of America (BofA) — BVA = US\$1.70tn, Citigroup (CITI) — BVA = US\$1.32tn, Wells Fargo (WELLS) — BVA = US\$1.24tn and Pittsburg National Corporation (PNC) — BVA = US\$290bn. As of Q4 2008 the AT200 represented US\$10.8tn in BVA, and of this the top five banks represent US\$6.4tn, or 59.4 per cent of the total. The skew in these data is extreme, as the average (median) bank among the top 200 has US\$53.8bn (US\$7.04bn) in BVA, reflected in a skewness coefficient of 6.8 that indicates a very elongated right tail relative to a normal distribution. Indeed, the top five banks reside well into the upper fifth percentile of the distribution of book value of assets (BVA = US\$162.9bn). The distribution of the book value of equity (BVE) is similarly skewed towards the largest banks, as the top 200(5) have aggregate BVE = US\$1.01tn (= US\$563.8bn, or 56.0 per cent of the top 200), as compared to the average (median) bank having MVE = US\$5.04bn (= US\$70m). It is seen that the distribution of the book value of total debt (BVTD) is even more extremely skewed towards the top five banks, the top 200(5) having BVTD = US\$9.75tn (= US\$5.83tn, or 60 per cent of the top 200), as compared to the average (median) bank having BVTD = US\$8.1bn.

Tables 2 and 3 summarise distributional properties of and correlations among the five accounting-based proxies for corresponding risk types. These are calculated from quarterly call reports in the period Q1 1984–Q4 2008, for the AT200 and top five banks. Credit risk (CR) is measured as *gross charge-offs* (GCO). Operational risk (OR) is measured as *other non-interest expense* (ONIE); for alternative approaches see Chernobai *et al.*[34] or De Fountnouvelle *et al.*[35] Market risk (MR) is proxied for by the deviation to the trailing

Table 1: Summary statistics on characteristics of top 200 and five largest banks by asset size (call report data as of 2008[1])

	Book value of total assets	Book equity	Book value of total debt	Book leverage ratio[2]	Lending assets	Percentage lending assets	Trading assets	Percentage trading assets	Total chargeoffs
Aggregate top 200 banks	10,758.51	1,007.19	9,751.33	90.64%	5,737.07	53.33%	964.24	8.96%	88.01
JP Morgan Chase	1,849.65	152.69	1,696.96	91.74%	738.44	39.92%	365.71	19.77%	10.75
Bank of America	1,699.71	178.72	1,520.99	89.49%	900.99	53.01 %	155.64	9.16%	17.60
Citigroup	1,319.45	101.46	1,217.99	92.31%	620.12	47.00%	200.52	15.20%	15.55
Wells Fargo	1,236.36	105.62	1,130.74	91.46%	792.49	64.10%	52.08	4.21%	7.52
PNC	289.88	25.25	264.62	91.29%	180.79	62.37%	6.09	2.10%	0.62
5th percentile	2.87	0.24	2.40	83.65%	1.94	39.64%	0.00	0.00%	0.0016
25th percentile	3.90	0.38	3.44	88.20%	2.66	62.06%	0.00	0.00%	0.01
Average bank	53.79	5.04	48.06	89.35%	28.69	66.63%	4.82	1.38%	0.44
Median bank	7.04	0.70	6.35	90.09%	4.38	69.53%	0.00	0.00%	0.04
75th percentile	15.47	1.65	14.21	91.81%	10.34	75.47%	0.05	0.40%	0.14
95th percentile	162.91	15.36	152.86	93.83%	92.30	86.12%	6.65	5.25%	1.38
Standard deviation	218.78	19.64	10.88	4.97%	109.41	15.40%	31.88	5.46%	1.92
Skewness	6.76	6.97	7.23	2.46	6.52	−1.45	9.13	7.28	7.20
Kurtosis	47.29	51.79	48.53	2.99	43.83	3.18	92.04	61.85	55.54

Continued

Table 1: Continued

	Chargeoff ratio[3]	Net interest income	Net-interest margin	Non-performing assets	Non-performing assets ratio[4]	Trading revenue	Non-interest income	Non-interest expense	Other non-interest expense
Aggregate top 200 banks	1.53%	289.33	5.04%	188.15	3.28%	−0.99	189.28	298.53	116.37
JP Morgan Chase	1.46%	43.38	5.88%	29.23	3.96%	5.02	41.78	46.35	14.40
Bank of America	1.95%	46.35	5.14%	27.82	3.09%	−0.35	29.94	36.13	13.67
Citigroup	2.51 %	35.40	5.71%	28.67	4.62%	−4.49	12.28	38.97	17.05
Wells Fargo	0.95%	35.40	4.47%	25.50	3.22%	0.35	22.16	32.89	12.21
PNC	0.34%	7.22	4.00%	29.23	16.17%	−0.13	3.35	9.38	2.85
5th percentile	0.04%	0.07	0.67%	0.01	0.36%	−0.07	0.00	0.00	0.00
25th percentile	0.36%	0.12	0.94%	0.05	1.18%	0.00	0.01	0.00	0.01
Average bank	1.27%	1.45	1.24%	0.94	3.40%	0.00	0.17	0.65	0.16
Median bank	0.76%	0.20	1.11%	0.11	2.37%	0.00	0.01	0.02	0.02
75th percentile	1.38%	0.48	1.31%	0.35	3.94%	0.00	0.06	0.10	0.04
95th percentile	3.93%	3.98	1.85%	2.06	9.75%	0.06	0.70	1.11	0.69
Standard deviation	1.95%	5.86	0.82%	3.94	4.60%	0.54	0.81	7.99	0.64
Skewness	6.44	6.59	7.16	6.52	5.61	0.81	8.09	22.03	6.99
Kurtosis	60.24	44.00	64.32	42.57	45.17	64.67	79.00	513.64	53.64

[1] Dollar amounts expressed in billions.
[2] Defined as the ratio of the book value of total debt to the book value of total assets.
[3] Defined as the ratio of gross charegeoffs to total lending assets.
[4] Defined as the ratio of non-performing assets to total lending assets.

Table 2: Summary statistics on risk measures for top 200 and five largest banks by asset size (call report data 1984–2008)[1]

		Min	5th prcntl.	25th prcntl.	Mean	Median	75th prcntl.	95th prcntl.	Max	Std. dev.	Skew.	Kurt.
Top 200 banks	Gross chargeoffs[2]	1.92	3.29	5.24	7.89	6.66	9.88	13.60	31.16	4.51	2.4931	9.0434
	Non-interest expense[3]	6.20	8.55	12.85	18.47	17.55	24.20	28.64	33.10	6.60	0.1041	−0.9597
	Net trading revenue[4]	−7.20	−1.79	−0.62	0.01	−0.13	0.25	1.30	16.13	2.14	4.5929	35.3144
	Liquidity gap[5]	−159.68	−112.11	−66.48	−20.10	−20.50	26.51	75.92	375.83	72.07	1.5928	7.8773
	Interest rate gap[6]	−171.72	−89.86	−57.76	−2.62	7.34	59.23	85.11	153.01	64.80	−0.0682	−0.6063
JPMC	Gross chargeoffs	0.79	1.07	1.38	1.96	1.74	2.32	3.59	4.53	0.82	1.2623	1.3433
	Non-interest expense	2.15	2.80	3.47	4.29	4.08	5.01	5.96	7.00	1.06	0.3278	−0.3711
	Net trading revenue	−1.59	−0.96	−0.32	−0.03	−0.09	0.21	0.87	3.65	0.63	1.9651	11.1842
	Liquidity gap	−82.87	−26.97	−12.15	2.28	−0.76	13.25	38.46	88.94	24.92	0.4594	2.2183
	Interest rate gap	−30.86	−16.66	−8.67	−0.13	0.92	8.46	14.76	25.60	10.92	−0.1630	−0.3518
Bank of America	Gross chargeoffs	0.85	1.21	1.52	2.10	1.91	2.43	3.91	5.81	0.87	1.6790	3.8429
	Non-interest expense	2.81	3.19	3.71	4.14	4.05	4.53	5.20	5.98	0.62	0.3662	0.1090
	Net trading revenue	−2.43	−0.36	−0.08	0.00	−0.01	0.05	0.25	4.38	0.61	2.8475	31.4616
	Liquidity gap	−65.05	−43.17	−20.41	−6.63	−6.85	5.92	35.65	84.95	24.59	0.5063	1.3789
	Interest rate gap	−34.99	−17.06	−9.37	−0.42	0.54	9.03	15.35	31.71	12.12	−0.1040	−0.0776
Citigroup	Gross chargeoffs	0.17	0.48	0.91	1.58	1.11	2.19	3.55	4.96	1.01	1.0753	0.2524
	Non-interest expense	0.69	1.67	2.21	3.04	2.58	3.72	5.21	6.43	1.20	0.7202	−0.1864
	Net trading revenue	−3.87	−0.39	−0.16	0.03	−0.02	0.06	0.43	9.09	1.04	6.2887	60.3677
	Liquidity gap	−34.38	−20.14	−10.08	−0.12	−1.93	7.55	20.76	90.05	17.96	2.2629	9.0858
	Interest rate gap	−15.34	−11.93	−6.42	−0.58	−0.91	5.83	11.39	14.60	7.32	−0.0244	−0.9156
Wells Fargo	Gross chargeoffs	0.45	0.60	0.83	1.15	1.00	1.33	2.21	3.50	0.52	1.7393	4.0686
	Non-interest expense	1.94	2.30	2.69	2.99	2.90	3.23	4.03	6.63	0.59	2.6639	14.1990
	Net trading revenue	−0.56	−0.12	−0.03	0.00	−0.01	0.02	0.12	0.75	0.13	2.0058	16.6127
	Liquidity gap	−41.58	−29.25	−12.26	−5.28	−4.91	0.29	19.90	30.69	13.30	0.0705	0.5398
	Interest rate gap	−24.80	−12.17	−6.31	−0.24	0.86	6.97	11.49	22.39	8.48	−0.1075	0.0474
PNC	Gross chargeoffs	0.10	0.17	0.21	0.33	0.27	0.38	0.68	1.33	0.20	2.4750	7.7287
	Non-interest expense	0.42	0.57	0.75	0.84	0.84	0.93	1.07	1.23	0.15	0.0660	0.4106
	Net trading revenue	−0.21	−0.03	−0.01	0.00	0.00	0.00	0.04	0.21	0.04	0.2049	22.0198
	Liquidity gap	−17.36	−10.17	−5.46	−1.59	−1.85	1.78	8.07	23.62	6.35	0.7620	2.3140
	Interest rate gap	−7.58	−3.76	−1.98	−0.02	0.58	2.43	3.52	7.26	2.80	−0.1077	−0.3254

[1]Dollar amounts expressed in billions.
[2]Gross chargeoffs (GCO) is the proxy measure credit risk (CR).
[3]Other non-interest expense (ONIE) is the proxy measure of operational risk (OR).
[4]The deviation to the trailing 4-quarter average in net-trading revenues (NTR-4QD) is the proxy measure of market risk (MR).
[5]The deviation to the trailing 4-quarter average of the liquidity gap, defined as total loans minus total deposits, the proxy measure of liquidity risk (LG-4QD).
[6]The deviation to the trailing 4-quarter average of the interest rate gap, defined as total interest expense minus total interest income, the proxy measure of interest rate risk (IRG-4QD).

Table 3: Pairwise correlations for top 200 and five largest banks risk proxies (call report data 1984–2008)

Risk pair	Type of correlation	Aggregate banks[2]	JP Morgan Chase	Bank of America	Citigroup	Wells Fargo	PNC
Credit and operational risk	Pearson	65.17%	−5.77%	−4.34%	76.65%	10.07%	28.87%
	Spearman	60.00%	−3.60%	−10.00%	78.00%	15.00%	41.00%
Credit and market risk	Pearson	22.41%	19.73%	5.29%	16.40%	18.42%	9.00%
	Spearman	−4.90%	15.00%	6.90%	8.10%	19.00%	9.00%
Credit and liquidity risk	Pearson	53.43%	19.07%	47.87%	31.47%	2.30%	20.85%
	Spearman	10.00%	−12.00%	−17.00%	−3.30%	−15.00%	−15.00%
Credit and interest rate risk	Pearson	−13.28%	−7.82%	−18.09%	−8.78%	−14.31%	−13.13%
	Spearman	33.00%	20.00%	24.00%	33.00%	17.00%	28.00%
Operational and market risk	Pearson	19.89%	10.92%	12.01%	13.46%	−4.28%	−9.31%
	Spearman	3.00%	10.00%	10.00%	2.70%	1.40%	−6.50%
Operational and liquidity risk	Pearson	15.33%	7.37%	−8.55%	11.76%	−4.85%	−10.22%
	Spearman	−2.00%	−16.00%	−24.00%	−9.20%	−26.00%	−18.00%
Operational and interest rate risk	Pearson	−11.74%	−14.25%	−23.49%	−8.79%	−15.88%	−15.68%
	Spearman	7.20%	10.00%	−30.00%	12.00%	−4.60%	−4.20%
Market and liquidity risk	Pearson	11.27%	1.56%	−18.23%	6.29%	−0.94%	−3.21%
	Spearman	2.30%	−36.00%	−23.00%	−23.00%	−25.00%	0.26%
Market and interest rate risk	Pearson	24.78%	−27.92%	−16.70%	−19.17%	−17.79%	3.38%
	Spearman	19.00%	−9.10%	8.80%	−0.60%	6.80%	3.90%
Interest rate and liquidity risk	Pearson	18.97%	19.96%	9.17%	12.38%	9.14%	12.86%
	Spearman	13.00%	21.00%	15.00%	26.00%	8.20%	18.00%

four-quarter average in *net-trading revenues* (NTR-4QD); such a measure is discussed in Jorion's book.[13] Whereas the proxy to CR of GCO is the same as in the paper of Rosenberg and Schuermann,[36] there is deviation from that in estimating OR and MR, for which the authors used external operational risk data and a GARCH factor model fit to macro data, respectively. In the extension of capturing liquidity risk (LR) and interest rate (or income) risk (IR), Jorion's prescriptions are also followed.[13] LR is approximated by the *liquidity gap*, defined as total loans minus total deposits, as a deviation from a moving four-quarter trailing average (LG-4QD). Similarly, IR is approximated by the *interest rate gap*, defined as total interest expense minus total interest income, as a deviation from a moving four-quarter trailing average (IRG-4QD).

Given these distributional features, when one implements the copula models, one chooses to model the marginal distributions of GCO and ONIE as two-parameter generalised extreme value (GEV) distributions, having non-negative support; and those of the remaining risk proxies (NTR-4QD, LG-4QD and IRG-4QD) as Student's t-distributions,[37] symmetric and with degrees of freedom determined by the data. One could fit alternative marginal distributions, potentially giving a better fit to the empirical distributions or better modelling to the tails. A relatively straightforward choice would be to use fitted kernel density estimators, at a modest but material increase in computational burden. An even more computationally expensive approach would be to model the body and the tails separately, say through a *conventional* distribution (eg lognormal or Student's t) and something like a generalised Pareto distribution, respectively. Nevertheless, one wishes to make the simplest parametric choices possible that are still conservative, in that these exhibit heavy tails relative to normal or log-normal. This is also for the purpose of making this exercise easily replicable by practitioners.

In Table 3 the linear Pearson (rank-order Spearman) correlations among the five proxies of the risk types are shown. It should be observed that these are the ordinary Pearson correlations among the rank-transformed variables. First, some wide disparities are seen across banks in the signs and magnitudes of the correlations. The second general observation is that magnitudes are generally on the low side, and in some cases negative, which would support the presence of substantial diversification benefits. Finally, the Spearman rank-order correlations also exhibit wide disparity in signs and magnitudes across risk pairs, and moreover are *not* generally in line with the results of the linear correlation analysis. The signs and magnitudes of the correlations are unevenly in line with empirical or theoretical evidence.[38–40]

The main results of this paper are tabulated in Tables 4, 5 and 6. In Table 4 are reported the 99.97th percentile VaR for alternative risk aggregation methodologies for each AT200 and the top five in row-wise panels. It is recognised that there are concerns about the coherence of VaR as a risk measure,[41,42] but results are found to be robust to using expected shortfall as the EC risk measure. The different techniques are arrayed by column as 'Gaussian copula simulation' (GCS), 'Gaussian (variance-covariance) approximation' (VCA), 'historical bootstrap (empirical copula) simulation' (ECS), 't-copula simulation' (TCS), 'Archimedean (Gumbel) copula simulation' (AGCS),[43] 'Archimedean copula (Clayton) simulation' (ACCS)[44] and 'Archimedean (Frank) copula simulation' (AFCS). The first row labelled 'magnitude of risk — fully diversified' represents the 99.97th percentile of the loss distribution, either simulated in the case of the copula methods or analytic in the normal approximation. The second rows of each panel labelled 'magnitude of risk — perfect correlation' represent the simple sum of the 99.97th percentiles of the simulated loss distributions for each risk type in the case of the copula methods, or $\Phi^{-1}(0.9997) = 3.43$ times the standard deviation of the total loss in the analytic normal approximation (in either case, referred to as 'simple summation of risks'). In the corresponding third rows is shown the 'proportional diversification benefit' (henceforth PDB), which is defined as the difference in the risk measure between the perfect correlation and fully diversified cases, expressed as a proportion of fully diversified VaR or ES

Table 4: 99.97% confidence level value-at-risk for five risk types: Credit, operational, market, liquidity and interest rate (200 largest banks: Call report data 1984–2008)

		Gaussian copula simulation	Gaussian (variance-covariance) approximation	Historical bootstrap (empirical copula) simulation	T-distribution copula simulation	Archimedean copula (Gumbel) simulation	Archimedean copula (Clayton) simulation	Archimedean copula (Frank) simulation
Top 200 banks	Magnitude of risk measure — Fully diversified	7.64E + 08	6.88E + 08	8.59E + 08	8.12E + 08	9.30E + 08	7.28E + 08	7.52E + 08
	Magnitude of risk measure — Perfect correlation	1.09E + 09	9.61E + 08	1.27E + 09	1.07E + 09	1.06E + 09	1.07E + 09	1.06E + 09
	Diversification benefit	42.05%	39.60%	48.17%	31.47%	14.51%	46.38%	41.69%
	Genest goodess of fit test P-value	3.55%	N/A	N/A	6.54%	0.05%	0.25%	1.15%
	VaR as a proportion of book value of assets	7.10%	6.40%	7.98%	7.55%	8.64%	6.76%	6.99%
JPMC	Magnitude of risk measure — Fully diversified	2.30E + 08	1.87E + 08	3.92E + 08	2.38E + 08	2.47E + 08	2.19E + 08	2.32E + 08
	Magnitude of risk measure — Perfect correlation	2.95E + 08	2.45E + 08	5.87E + 08	3.16E + 08	2.97E + 08	2.97E + 08	2.96E + 08
	Diversification benefit	28.27%	31.17%	49.94%	32.60%	20.18%	35.42%	27.74%
	Genest goodess of fit test P-value	20.53%	N/A	N/A	7.24%	0.05%	61.69%	37.23%
	VaR as a proportion of book value of assets	12.43%	10.12%	21.18%	12.88%	13.35%	11.85%	12.54%
Bank of America	Magnitude of risk measure — Fully diversified	1.94E + 08	1.82E + 08	2.05E + 08	2.00E + 08	2.07E + 08	1.82E + 08	2.03E + 08
	Magnitude of risk measure — Perfect correlation	2.69E + 08	2.48E + 08	2.89E + 08	2.75E + 08	2.60E + 08	2.65E + 08	2.64E + 08
	Diversification benefit	38.31%	36.32%	40.93%	37.50%	25.22%	45.34%	30.13%
	Genest goodess of fit test P-value	50.10%	N/A	N/A	6.04%	42.91%	60.79%	39.01%
	VaR as a proportion of book value of assets	11.43%	10.72%	12.05%	11.75%	12.19%	10.72%	11.93%

Continued

Table 4: Continued

		Gaussian copula simulation	Gaussian (variance-covariance) approximation	Historical bootstrap (empirical copula) simulation	T-distribution copula simulation	Archimedean copula (Gumbel) simulation	Archimedean copula (Clayton) simulation	Archimedean copula (Frank) simulation
Citigroup	Magnitude of risk measure — Fully diversified	1.62E + 08	1.32E + 08	2.77E + 08	1.72E + 08	2.00E + 08	1.49E + 08	1.60E + 08
	Magnitude of risk measure — Perfect correlation	2.19E + 08	1.83E + 08	3.92E + 08	2.20E + 08	2.23E + 08	2.18E + 08	2.18E + 08
	Diversification benefit	35.23%	37.91%	41.42%	27.53%	11.22%	46.05%	36.41%
	Genest goodess of fit test P-value	32.82%	N/A	N/A	6.04%	30.12%	12.04%	6.94%
	VaR as a proportion of book value of assets	12.28%	10.03%	21.03%	13.07%	15.19%	11.29%	12.13%
Wells Fargo	Magnitude of risk measure — Fully diversified	1.63E + 08	1.04E + 08	1.87E + 08	1.71E + 08	1.99E + 08	1.52E + 08	1.58E + 08
	Magnitude of risk measure — Perfect correlation	2.18E + 08	1.47E + 08	2.78E + 08	2.19E + 08	2.20E + 08	2.20E + 08	2.21E + 08
	Diversification benefit	33.97%	41.03%	49.07%	27.82%	10.35%	44.76%	39.28%
	Genest goodess of fit test P-value	28.72%	N/A	N/A	57.19%	30.02%	8.44%	8.14%
	VaR as a proportion of book value of assets	13.18%	8.45%	15.11%	13.86%	16.11%	12.28%	12.81%
PNC	Magnitude of risk measure — Fully diversified	4.79E + 07	4.66E + 07	5.78E + 07	5.01E + 07	5.23E + 07	4.66E + 07	4.71E + 07
	Magnitude of risk measure — Perfect correlation	6.33E + 07	6.10E + 07	7.92E + 07	6.78E + 07	6.33E + 07	6.35E + 07	6.41E + 07
	Diversification benefit	32.16%	31.12%	37.04%	35.21%	21.09%	36.39%	36.09%
	Genest goodess of fit test P-value	47.70%	N/A	N/A	35.35%	82.07%	10.94%	6.51%
	VaR as a proportion of book value of assets	16.52%	16.06%	19.94%	17.30%	18.04%	16.06%	16.25%

Table 5: Bootstrap analysis of 99.97% confidence level value-at-risk for five risk types: Credit, operational, market, liquidity and interest rate (200 largest banks: Call report data 1984–2008)

			Resampling of correlations					Resampling of margins			
		Empirical copula	Normal approx.	Gaussian copula	T-copula	Gumbel copula	Clayton copula	Gaussian copula	T-copula	Gumbel copula	Clayton copula
Top 200 banks	99.97% VaR in historical sample	8.59E + 08	6.88E + 08	7.64E + 08	8.12E + 08	7.28E + 08	7.28E + 08	7.64E + 08	8.12E + 08	7.28E + 08	7.28E + 08
	95% confidence interval	6.96E + 07	1.94E + 08	5.41E + 07	6.90E + 07	4.77E + 07	4.77E + 07	2.70E + 08	3.54E + 08	1.05E + 09	8.68E + 08
	Numerical coefficient of variation	8.10%	28.20%	7.08%	8.49%	6.55%	6.55%	35.37%	43.61%	39.32%	44.29%
JPMC	99.97% VaR in historical sample	3.92E + 08	1.87E + 08	2.30E + 08	2.38E + 08	2.47E + 08	2.19E + 08	2.30E + 08	2.38E + 08	2.47E + 08	2.19E + 08
	95% confidence interval	2.52E + 07	5.68E + 07	1.79E + 07	1.79E + 07	1.86E + 07	1.53E + 07	9.52E + 07	1.11E + 08	8.28E + 07	1.09E + 08
	Numerical coefficient of variation	6.42%	30.34%	7.78%	7.51%	7.53%	7.00%	41.38%	46.64%	33.51%	49.89%
Bank of America	99.97% VaR in historical sample	2.05E + 08	1.82E + 08	1.94E + 08	2.00E + 08	2.07E + 08	1.82E + 08	1.94E + 08	2.00E + 08	2.07E + 08	1.82E + 08
	95% confidence interval	2.78E + 07	5.07E + 07	1.72E + 07	2.21E + 07	2.24E + 07	1.18E + 07	9.37E + 07	1.12E + 08	7.83E + 07	4.58E + 07
	Numerical coefficient of variation	13.57%	27.86%	8.85%	11.09%	10.80%	6.50%	48.21%	56.14%	37.79%	25.16%
Citigroup	99.97% VaR in historical sample	2.77E + 08	1.32E + 08	1.62E + 08	1.72E + 08	2.00E + 08	1.49E + 08	1.62E + 08	1.72E + 08	2.00E + 08	1.49E + 08
	95% confidence interval	2.52E + 07	5.99E + 07	1.29E + 07	2.07E + 07	1.92E + 07	9.71E + 06	6.22E + 07	1.07E + 08	1.06E + 08	8.10E + 07
	Numerical coefficient of variation	9.08%	45.26%	7.95%	12.00%	9.57%	6.52%	38.41%	62.23%	52.66%	54.38%

Continued

Table 5: Continued

		Empirical copula	Normal approx.	Resampling of correlations				Resampling of margins			
				Gaussian copula	T-copula	Gumbel copula	Clayton copula	Gaussian copula	T-copula	Gumbel copula	Clayton copula
Wells Fargo	99.97% VaR in historical sample	1.87E + 08	1.04E + 08	1.63E + 08	1.71E + 08	1.99E + 08	1.52E + 08	1.63E + 08	1.71E + 08	1.99E + 08	1.52E + 08
	95% confidence interval	7.39E + 06	2.82E + 07	1.32E + 07	1.66E + 07	1.84E + 07	9.12E + 06	6.52E + 07	7.88E + 07	9.11E + 07	9.44E + 07
	Numerical coefficient of variation	12.78%	26.98%	8.10%	8.91%	9.25%	6.00%	40.00%	45.99%	45.74%	62.14%
PNC	99.97% VaR in historical sample	5.78E + 07	4.66E + 07	4.79E + 07	5.01E + 07	5.23E + 07	4.66E + 07	4.79E + 07	5.01E + 07	5.23E + 07	4.66E + 07
	95% confidence interval	7.39E + 06	1.53E + 07	4.29E + 06	5.46E + 06	5.22E + 06	2.75E + 06	2.15E + 07	2.38E + 07	2.35E + 07	3.24E + 07
	Numerical coefficient of variation	12.78%	32.85%	8.96%	10.89%	9.98%	5.90%	44.82%	47.44%	44.94%	69.55%

Table 6: Bootstrap analysis 99.97 per cent confidence percent level diversification benefit for five risk types: Credit, operational, market, liquidity and interest rate (200 largest banks: Call report data 1984–2008)

			Resampling of correlations					Resampling of margins			
		Empirical copula	Normal approx.	Gaussian copula	T-copula	Gumbel copula	Clayton copula	Gaussian copula	T-copula	Gumbel copula	Clayton copula
Top 200 banks	Diversification % in historical sample	4.82E − 01	3.96E − 01	4.21E − 01	3.15E − 01	1.45E − 01	4.64E − 01	4.21E − 01	3.15E − 01	1.45E − 01	4.64E − 01
	95% confidence interval	6.85E − 02	4.44E − 01	9.88E − 02	9.85E − 02	1.10E − 01	1.11E − 01	1.95E − 01	1.93E − 01	5.34E − 02	1.62E − 01
	Numerical coefficient of variation	14.22%	112.15%	23.48%	31.31%	75.46%	23.89%	46.38%	61.37%	36.82%	35.01%
JPMC	Diversification % in historical sample	4.99E − 01	3.12E − 01	2.83E − 01	3.26E − 01	2.02E − 01	3.54E − 01	2.83E − 01	3.26E − 01	2.02E − 01	3.54E − 01
	95% confidence interval	6.20E − 02	3.47E − 01	9.70E − 02	1.17E − 01	8.91E − 02	1.06E − 01	1.57E − 01	2.26E − 01	7.84E − 02	1.16E − 01
	Numerical coefficient of variation	12.41%	111.16%	34.32%	35.99%	44.12%	30.00%	55.64%	69.44%	38.83%	32.84%
Bank of America	Diversification % in historical sample	4.09E − 01	3.63E − 01	3.83E − 01	3.75E − 01	2.52E − 01	4.53E − 01	3.83E − 01	3.75E − 01	2.52E − 01	4.53E − 01
	95% confidence interval	7.40E − 02	3.64E − 01	9.36E − 02	1.48E − 01	1.18E − 01	1.13E − 01	1.48E − 01	1.48E − 01	1.12E − 01	1.20E − 01
	Numerical coefficient of variation	18.08%	100.26%	24.44%	39.46%	46.76%	24.93%	38.63%	39.46%	44.47%	26.54%
Citigroup	Diversification % in historical sample	4.14E − 01	3.79E − 01	3.52E − 01	2.75E − 01	1.12E − 01	4.60E − 01	3.52E − 01	2.75E − 01	1.12E − 01	4.60E − 01
	95% confidence interval	5.16E − 02	3.73E − 01	1.02E − 01	1.17E − 01	9.93E − 02	1.12E − 01	1.59E − 01	2.31E − 01	5.46E − 02	2.01E − 01
	Numerical coefficient of variation	12.46%	98.43%	29.02%	42.37%	88.45%	24.24%	45.08%	83.94%	48.61%	43.72%
Wells Fargo	Diversification % in historical sample	4.91E − 01	4.10E − 01	3.40E − 01	2.78E − 01	1.04E − 01	4.48E − 01	3.40E − 01	2.78E − 01	1.04E − 01	4.48E − 01
	95% confidence interval	4.87E − 02	3.42E − 01	5.88E − 02	7.89E − 02	9.10E − 02	1.02E − 01	1.49E − 01	1.62E − 01	5.50E − 02	2.21E − 01
	Numerical coefficient of variation	9.92%	83.43%	17.30%	28.35%	87.93%	22.73%	43.85%	58.20%	53.09%	49.48%

Continued

Table 6: Continued

			Resampling of correlations					Resampling of margins			
		Empirical copula	Normal approx.	Gaussian copula	T-copula	Gumbel copula	Clayton copula	Gaussian copula	T-copula	Gumbel copula	Clayton copula
PNC	Diversification % in historical sample	3.70E − 01	3.11E − 01	3.22E − 01	3.52E − 01	2.11E − 01	3.64E − 01	3.22E − 01	3.52E − 01	2.11E − 01	3.64E − 01
	95% confidence interval	6.70E − 02	3.68E − 01	7.88E − 02	7.91E − 02	1.14E − 01	1.03E − 01	1.82E − 01	2.14E − 01	9.08E − 02	7.25E − 02
	Numerical coefficient of variation	18.10%	118.15%	24.50%	22.45%	53.88%	28.27%	56.68%	60.91%	43.03%	19.93%

for the respective tables:

$$\% \, Diversification \, Benefit^{\, VaR_{99.97\%}} = \frac{VaR^{VaR_{99.97\%}}_{Perfect Correlation} - VaR^{VaR_{99.97\%}}_{Fully \, Diversified}}{VaR^{VaR_{99.97\%}}_{Fully \, Diversified}} \quad (3.1)$$

$$\% \, Diversification \, Benefit^{\, ES_{99.9\%}} = \frac{VaR^{ES_{99.9\%}}_{Perfect Correlation} - VaR^{ES_{99.9\%}}_{Fully \, Diversified}}{VaR^{ES_{99.9\%}}_{Fully \, Diversified}} \quad (3.2)$$

In the second-to-bottom rows are tabulated the p-values of recently developed goodness-of-fit tests for the copula models.[45,46] Approximate p-values for this test are based upon a comparison of the ECS to a parametric estimate of the copula in question, that is generated through a parametric bootstrap, under the null hypothesis the data are generated through the ECS process.

In the bottom row is shown the diversified VaR as a proportion of the BVTA, for AT200 and each top five institution as of the year-end 2008. Wide variation is observed in all risk and diversification measures across aggregation methodologies for a given institution, as well as across banks for a given technique. The dollar VaR is increasing in size of the institution, ranging across diversification methodologies: US\$688bn–US\$930bn for AT200, US\$187bn–US\$392bn for JPMC, US\$182bn–US\$207bn for BofA, US\$132bn–US\$277bn for CITI, US\$104bn–US\$199bn for WELLS, and finally a big drop-off US\$46.6bn–US\$57.8bn for PNC. VaR expressed as a proportion of BVA also shows much variation across both aggregation techniques and institutions, ranging from the middle single-digits to just below 20 per cent. These percentages generally decrease with the size of the institution, although the relationship is not strictly monotonic. The lowest percentages are observed for the hypothetical aggregate AT200 (6 per cent–9 per cent), highest for PNC (16 per cent–18 per cent), and generally hovering just north of 10 per cent for the middle four banks: 10 per cent–20 per cent, 11 per cent–12 per cent, 10 per cent–15 per cent and 8 per cent–16 per cent for JPMC, BofA, CITI and WELLS, respectively. Note that the ranges of VaR/BVA across methodologies appear to be increasing from JPMC down to WELLS. There is caution in concluding much from this, such as a ‘business line diversification story’, owing to the small sample size.

Comparing different risk aggregation methodologies across banks, it is observed that VCA produces consistently the lowest VaR, and that either the ECS or the AGCS produces the highest VaR, across all institutions. ECS and ACGS are followed by TCS in terms of conservativeness, while the GCS ‘benchmark’ is usually somewhere in the middle, and ACCS is towards the low side. AFGS tends to be closest to GCS, albeit usually just a little lower. While TCS is always higher than GCS, in some cases it is not by a very wide margin.

In the case of AT200, VaR under ECS (VCA) is US\$859bn (US\$688bn), US\$392bn (US\$187bn), US\$205bn (US\$182bn), US\$277bn (US\$132bn), US\$187bn (US\$104bn) and US\$57.8bn (US\$46.6bn) for AT200, JPMC, BofA, CITI, WELLS and PNC, respectively; and this brackets the respective GCS VaRs of US\$764bn, US\$230bn, US$194bn, US$162bn, US$163bn and US$47.9bn. AGCS is in some cases close to ECS, and in others still higher than GCS (understandably, with the property of *upper tail dependence*), with VaRs of US$930bn, US$247bn, US$207bn, US$200bn, US$199bn and US$52.3bn for AT200, JPMC, BofA, CITI, WELLS and PNC, respectively. On the other hand, TCS is always higher than GCS, but in some cases by only a modest amount (and generally less than AGCS or ECS): VaRs of US$812bn, US$238bn, US$200bn, US$172bn, US$171bn and US$50.1bn for AT200, JPMC, BofA, CITI, WELLS and PNC, respectively. The ACCS is generally second place to VCA in lack of conservativeness, understandably so given its property of *lower tail dependence*: VaRs of US$728bn, US$219bn, US$182bn, US$149bn, US$152bn and US$46.6bn for AT200, JPMC, BofA, CITI, WELLS and PNC, respectively. Finally, it is seen that the AFCS (the Archimedean copula characterised by *neither upper nor lower tail dependence*) is middling and often close to GCS in VaR magnitude as compared to its brethren methodologies: VaRs of US$752bn, US$232bn, US$203bn, US$160bn, US$158bn and US$47.1bn for AT200, JPMC, BofA, CITI, WELLS and PNC, respectively.

The proportional diversification benefits, or PDBs, exhibit a great deal of variation across banks and aggregation techniques, ranging from 10 per cent to 50 per cent, with the ECS (AGCS) yielding clearly higher (lower) values than the other methodologies. PDBs of ECS range from 40 per cent–50 per cent, while they range from 10 per cent–25 per cent for AGCS. Across banks, the GCS 'benchmark' tends to lie in the middle (41 per cent–58 per cent), and the VCA to the lower end of the range (31 per cent–41 per cent), while AGCS is the lowest (10 per cent–21 per cent). Looking at the range of the PDBs across aggregation methodologies for a given bank, an attempt is made to measure the impact of business mix. Yet, a directionally consistent pattern cannot be observed. The two banks with the highest proportion of trading assets have diversification benefits spanning a low and wide (11.2 per cent–46.1 per cent for CITI) to a high and narrow range (20.3 per cent–36.6 per cent for JPMC). On the other hand, considering banks with proportionately more lending assets, BofA has a range similar to JPMC (25.2 per cent–38.3 per cent), while Wells and PNC more closely resemble CITI (10.4 per cent–49.1 per cent and 21.1 per cent–37.0 per cent, respectively). There is no discussion about the 99th percentile expected shortfall (ES) (table available on request), as *generally* the results are quite similar in both absolute quantities and in comparisons across institutions or aggregation methodologies.

The results of the GOF tests[46] are highly mixed (the null being rejected in just under one-half of cases, 14 out of 30) and do not lend themselves to the extraction of a clear pattern. Generally, the rejections of fit to the empirical processes are not at very high levels of significance, so that perhaps one can say that the models are doing a decent job. There are only three rejections at better than the 1 per cent level (AGCS for AT200 and JPMC, and AGCS for JPMC), only one at the 5 per cent level (AFCS for AT200), and the remaining nine at only the 10 per cent level (and in one case, the p-value is just above 0.10). AT200 has the most rejections (in all cases, models are rejected at the 10 per cent level), followed by JPMC (two rejections for TCS and AGCS), CITI and WELLS (two rejections each at the 10 per cent level), with BofA and PNC having the least (only one each at the 10 per cent level). Across banks, the GCS and AGCS models fail to reject a fit to the data most often (one and two rejections, respectively; however, AGCS has the second lowest p-value), while the TCS (four at the 10 per cent level) and AFCS (three at the 10 per cent level and one at the 5 per cent level) have the most rejections.

The object of the second analysis that is performed, a bootstrap (or resampling) exercise, is to measure the uncertainty in the VaR and PDB estimates. This is now a widely used technique in finance and economics, originating mainly in the statistics literature, which has the potential to develop estimates of standard errors or confidence intervals for complex functions of random variables for which distribution theory is undeveloped.[31] The type of bootstrap that is implemented is the so-called *non-parametric* version, in which the data are resampled with replacements. In each iteration, the function of interest is recalculated, yielding a distribution of the latter which can be analysed. As the VaR estimate in any of the aggregation frameworks depends upon a random sample of observations, and in the case of the VCA or the copulas parametric estimates of the marginal distributions or of the correlation matrices, the uncertainty in the latter flows through to the former. This manner of analysis is of keen importance to regulators, as they must seek to understand how one may decompose the volatility of capital from year to year into that driven by the variability in model inputs, distinguishing from that stemming from changes to a bank's risk profile.

The results of this experiment are tabulated in Tables 5 and 6 for the VaR and PDB estimates, respectively. One resamples with replacement 10,000 times, and in each bootstrap a simulation of 100,000 years is run as in the main results. The *numerical coefficients of variation* (NCV) of VaR and PDB across banks and techniques are shown in the final rows of each panel in Tables 5 and 6. The NCV is defined as the ratio of the 95 per cent confidence interval in the bootstrapped sample to the estimate in the historical sample:

$$NCV_{95\%}^{VaR_{99.97\%}} = \frac{VaR^{99.97\%}_{BTSTRP_{Q97.5\%}} - VaR^{99.97\%}_{BTSTRP_{Q2.5\%}}}{VaR^{99.97\%}_{SMPL}} \tag{3.3}$$

$$NCV_{95\%}^{PDB_{99.97\%}} = \frac{PDB^{99.97\%}_{BTSTRP_{Q97.5\%}} - PDB^{99.97\%}_{BTSTRP_{Q2.5\%}}}{PDB^{99.97\%}_{SMPL}} \tag{3.4}$$

This bootstrap is done in two ways: holding the estimates of the marginal distributions constant, and re-estimating the correlations, and vice versa (ie assuming that the true correlation matrix is known, but that the parameters of the marginal distribution are measured with statistical error), shown in the left and right panels of the tables, respectively. Yet, in the case of ESC, neither of these can be done and it will be necessary to draw a new sample from which an empirical copula from the resampled data can be estimated. And in the case of the VCA, only the correlation resampling can be done, as that methodology does not depend upon fitting marginal distributions.

There are several clear conclusions that can be drawn based upon these results. First, a consistent pattern in the variability of VaR or PDB across size or types of banks (ie business mix) is mixed. Secondly, regarding which model is most or least stable, it is observed that, for either the bootstrap of VaR or PDB, the ECS and GSC techniques yield generally the lowest NCVs as compared to other methodologies. Thirdly, in contrast to this, the VCA is consistently the most variable in the bootstrap, having for the most part the highest NCVs. In the comparison between VCA and the copula methods (excluding ECS) this is somewhat surprising, since VCA does not require estimation of marginal distribution parameters, yet nonetheless has much higher NCVs in the resampling of correlations for any of the copula methodologies. In the bootstrap of VaR, NCV ranges from 6.4 per cent–13.6 per cent for ECS and 27.9 per cent–45.3 per cent for VCA, while in the resampling of correlations (margins) for GCS they range from 7.1 per cent–9.0 per cent (35.4 per cent–48.2 per cent). Fourthly, for either the bootstrap of VaR or PDB, NCVs are an order of magnitude higher for the resampling of margins as compared to the resampling of correlations, and this difference is accentuated for the bootstrapping of VaR as compared to PDB. Fifthly, NCVs are higher for the PDB as compared to the VaR statistics. In the case of VaR, NCVs in the bootstrap of correlations (margins) range from 5.9 per cent–45.3 per cent (25.2 per cent–69 per cent), while in the case of PDB the corresponding numbers are 9.9 per cent–158.2 per cent (22.7 per cent–118.2 per cent). Finally, according to the NCV criterion, the PDB is much more imperfectly estimated than the VaR, across methodologies or banks.

As with the VaR bootstrapping, in the resampling of PDB in Table 6, it can be seen that overall NCVs are higher than in the estimation of VaR, across methodologies and institutions. The estimation of PDB is least precise for VCA, and generally most accurate for ECS, followed closely by GCS in having low NCVs. In the resampling of correlations, the GCS and ACCS are notably more variable, in that order, as compared with the VaR estimation. As with the case of VaR, the resampling of margins has higher NCVs, although the difference as compared to the correlation bootstrapping is not as stark as in VaR estimation.

CONCLUSION

In this study the state of practice in the validation of economic capital models has been surveyed. Several examples have been presented along the way, highlighting the differences, difficulties and ambiguities in the validation of this class of risk models. This contrasted with models in which the state of the art is more established, such as models of trading risk or default prediction; however, also highlighted were analogies to models in which there are similar issues, such as those for credit or operational VaR, or even PD models in low-default portfolio settings. After motivating the survey, a framework was presented for categorising and understanding the validation practices, proceeding from qualitative techniques deemed the most qualitative (eg use test standards and analysis of conceptual soundness) to the least qualitative (examinations of assumptions); followed by quantitative techniques, proceeding similarly from the least quantitative (sensitivity analysis) to the most quantitative (benchmarking, model replication and backtesting) of those techniques. Also discussed were supervisory expectation and concerns, which it is argued overlap with those of properly incentivised risk managers. It concluded with an example from a risk aggregation study using US bank's loss data, which illustrates many of these quantitative techniques (ie benchmarking of models), and found that several commonly used economic capital models compared unevenly with a lesser-known benchmark. This latter exercise highlights the importance on the part of banks of carefully considering the implications of potential flaws in their EC modelling frameworks for the purpose of managing model risk. It is hoped and believed that a sound contribution has been made to the industry's knowledge base on the topic of validating EC models.

Author's note

The views expressed herein are those of the author and do not necessarily represent a position taken either by the US Office of the Comptroller of the Currency or the US Department of the Treasury.

References

1 PricewaterhouseCoopers (2005) 'Effective Capital Management: Economic Capital as an Industry Standard', PricewaterhouseCoopers, New York, NY.
2 BCBS (2009) 'Range of Practices and Issues in Economic Capital Framework', BCBS, Basel, Switzerland.
3 Bank of Japan (2005) 'Advancing Integrated Risk Management', Tokyo, Japan.
4 Bank of Japan (2007) 'Economic Capital Workshop Summary Record', Tokyo, Japan.
5 BCBS (2004) 'International Convergence of Capital Measurement and Capital Standards: A Revised Framework', BCBS, Basel, Switzerland.
6 BCBS (2005) 'Update on Work of the Accord Implementation Group Related to Validation under the Basel II Framework', BCBS, Basel, Switzerland.
7 Embrechts, P., McNeil, A. and Straumann, D. (1999) 'Correlation: Pitfalls and alternatives', *Risk*, Vol. 12, pp. 69–71.
8 Embrechts, P., McNeil, A. and Straumann, D. (2002) 'Correlation and dependence in risk management: Properties and pitfalls', in Dempster, M. (ed.) 'Risk Management: Value at Risk and Beyond', Cambridge University Press, Cambridge, UK, pp. 176–223.
9 Embrechts, P., Lindskog, F. and McNeil, A. (2003) 'Modelling dependence with copulas and applications to risk management', in Rachev, S. (ed.) 'Handbook of Heavy Tailed Distributions in Finance', Elsevier, Rotterdam, pp. 329–384.
10 BCBS Joint Forum (2003) 'Trends in Risk Integration and Aggregation', BCBS, Basel, Switzerland.
11 Kuritzkes, A., Rosenberg, J., Schuermann, T. and Weiner, S. (2003) 'Risk measurement, risk management and capital adequacy of financial conglomerates', in Herring, R. and Litan, R. (eds), Brookings-Wharton Papers in Financial Services, Philadelphia, PA, pp. 141–194.
12 Rutter Associates LLC (2004) 'Rutter Associates Survey of Credit Portfolio Management Practices', Rutter Associates, New York, NY.
13 Inanoglu, H. and Jacobs, M. (2009) 'Models for risk aggregation and sensitivity analysis: An application to bank economic capital', *Journal of Risk and Financial Management*, Vol. 2, pp. 118–189.

14 Basel Committee on Banking Supervision (BCBS) (1988) 'International Convergence of Capital Measurement and Capital Standard', BCBS, Basel, Switzerland.
15 Phelan, M. J. (1995) 'Probability and statistics applied to the practice of financial risk management: The case of JP Morgan's RiskMetrics™', working paper, Wharton Financial Institutions Center, Philadelphia, PA, 95, pp. 1–19.
16 BCBS (1999) 'Credit Risk Modelling: Current Practices and Applications', BCBS, Basel, Switzerland.
17 BCBS Joint Forum (2001) 'Risk Management Practices and Regulatory Capital, Cross-sectoral Comparison', BCBS, Basel, Switzerland.
18 IFRI and CRO Forum (2007) 'Insights from the Joint IFRI/CRO Forum Survey on Economic Capital Practice and Applications', Switzerland.
19 IACPM and ISDA (2006) 'Convergence of Credit Capital Models', Rutter Associates LLC.
20 Burns, R. (2004) 'Economic capital and the assessment of capital adequacy', *Supervisory Insights*, Vol. 2, No. 2, pp. 5–11.
21 BCBS (2005) 'Studies on the Validation of Internal Rating Systems', BCBS, Basel, Switzerland.
22 Jacobs, M. and Kiefer, N. (2010) 'The Bayesian approach to default risk: A guide', forthcoming in Boecker, K. (ed.) 'Rethinking Risk Measurement and Reporting', Risk Books, London.
23 BCBS (2005) 'Validation of Low-default Portfolios in the Basel II Framework', BCBS, Basel, Switzerland.
24 Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2002) 'Portfolio value-at-risk with heavy-tailed risk factors', *Mathematical Finance*, Vol. 12, No. 3, pp. 239–270.
25 Jorion, P. (2006) 'Value at Risk: The Benchmark for Managing Financial Risk', 3rd edn, McGraw Hill, New York, NY.
26 Office of the Comptroller of the Currency (OCC)/Department of the Treasury, Federal Reserve System, Office of Thrift Supervision/Department of the Treasury (2007) 'Risk-Based Capital Standards: Advanced Capital Adequacy Framework — Basel II', Federal Register, Vol. 72, No. 235, 7th December ['Final Rule'], available at: http://www.occ.treas.gov/fr69288.pdf (accessed 16th August, 2010).
27 Sklar, A. (1956) 'Fonctions de répartition à n dimensions et leurs marges', Publications de l'Institut de Statistique de L'Université de Paris, 8, pp. 229–231.
28 Nelson, R. (1999) 'An Introduction to Copulas', Springer, New York, NY.
29 Li, D. (2000) 'On default correlation: A copula function approach', *Journal of Fixed Income*, Vol. 9, No. 4, pp. 43–54.
30 Gouriéroux, C., Laurent, J. and Scaillet, O. (2000) 'Sensitivity analysis of values at risk', *Journal of Empirical Finance*, Vol. 7, No. 3, pp. 225–245.
31 Efron, B. and Tibshirani, R. (1986) 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', *Statistical Science*, Vol. 1, No. 1, pp. 54–75.
32 Frey, R. and McNeil, A. (2001) 'Modelling dependent defaults', working paper, ETH E-collection, Zurich.
33 Fermanian, J. and Scaillet, O. (2003) 'Nonparametric estimation of copulas for time series', *Journal of Risk*, Vol. 5, No. 4, pp. 25–54.
34 Chernobai, A., Jorion, P. and Yu, F. (2008) 'The determinants of operational risk', working paper, University of Syracuse.
35 De Fountnouvelle, P., DeJesus-Rueff, V., Jordan, J. and Rosengren, E. (2003) 'Using loss data to quantify operational risk', working paper, Federal Reserve Bank of Boston, Boston, MA.

36 Rosenberg, J. and Schuermann, T. (2006) 'A general approach to integrated risk management with skewed, fat-tailed risks', *Journal of Financial Economics*, Vol. 79, No. 3, pp. 569–614.
37 Demarta, S. and McNeil, A. (2005) 'The T copula and related copulas', *International Statistical Review*, Vol. 73, No. 1, pp. 111–130.
38 Cherubini, U. and Lunga, G. (2001) 'Liquidity and credit risk', *Applied Mathematical Finance*, Vol. 8, No. 2, pp. 79–95.
39 Alexander, C. and Pézier, J. (2003) 'On the aggregation of firm-wide market and credit risks', Discussion papers in finance 2003–13, ICMA Centre, Reading, MA.
40 Ericsson, J. and Renault, O. (2002) 'Liquidity and credit risk', working paper, EFA 2003 Glasgow, available at: http://ssrn.com/abstract=424002 (accessed 16th August, 2010).
41 Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1997) 'Thinking coherently', *Risk*, Vol. 10, No. 11, pp. 68–71.
42 Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1999) 'Coherent measures of risk', *Mathematical Finance*, Vol. 9, No. 3, pp. 203–228.
43 Gumbel, E. (1960) 'Distributions des Valeurs Extremes en Plusieurs Dimensions', Publications de l'Institut de Statistique de l'Université de Paris, Vol. 9, pp. 171–173.
44 Clayton, D. (1978) 'A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence', *Biometrika*, Vol. 65, No. 1, pp. 141–151.
45 Genest, C. and Beaudoin, D. (2004) 'Tests of independence and randomness based on the empirical copula process', *Test*, Vol. 13, No. 2, pp. 335–369.
46 Genest, C. and Rémillard, B. (2009) 'Goodness of fit tests for copulas: A review and power study', *Insurance Mathematics and Economics*, Vol. 44, pp. 199–213.
47 BCBS (1996) 'Amendment to the Capital Accord to Incorporate Market Risks', BCBS, Basel, Switzerland.