Iason ltd. and Energisk.org are the editors of Argo newsletter. Iason is the publisher. No one is allowed to reproduce or transmit any part of this document in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of Iason ltd. Neither editor is responsible for any consequence directly or indirectly stemming from the use of any kind of adoption of the methods, models, and ideas appearing in the contributions contained in Argo newsletter, nor they assume any responsibility related to the appropriateness and/or truth of numbers, figures, and statements expressed by authors of those contributions.
Editors:
Antonio CASTAGNA (Co-founder of Iason ltd and CEO of Iason Italia srl)
Andrea RONCORONI (ESSEC Business School, Paris)

Executive Editor:
Luca OLIVO (Iason ltd)

Scientific Editorial Board:
Alvaro CARTEA (University College London)
Antonio CASTAGNA (Co-founder of Iason ltd and CEO of Iason Italia srl)
Mark CUMMINS (Dublin City University Business School)
Gianluca FUSAI (Cass Business School, London)
Sebastian JAIMUNGAL (University of Toronto)
Fabio MERCURIO (Bloomber LP)
Andrea RONCORONI (ESSEC Business School, Paris)

Iason ltd
Registered Address:
6 O'Curry Street
Limerick 4
Ireland

Italian Address:
Piazza 4 Novembre, 6
20124 Milano
Italy

Contact Information:
info@iasonltd.com
www.iasonltd.com

Energisk.org
Contact Information:
contact@energisk.org
www.energisk.org

Iason ltd and Energisk.org are registered trademark.

Articles submission guidelines

Argo welcomes the submission of articles on topical subjects related to the risk management. The two core sections are Banking and Finance and Energy and Commodity Finance. Within these two macro areas, articles can be indicatively, but not exhaustively, related to models and methodologies for market, credit, liquidity risk management, valuation of derivatives, asset management, trading strategies, statistical analysis of market data and technology in the financial industry. All articles should contain references to previous literature. The primary criteria for publishing a paper are its quality and importance to the field of finance, without undue regard to its technical difficulty. Argo is a single blind refereed magazine: articles are sent with author details to the Scientific Committee for peer review. The first editorial decision is rendered at the latest within 60 days after receipt of the submission. The author(s) may be requested to revise the article. The editors decide to reject or accept the submitted article. Submissions should be sent to the technical team (argo@iasonltd.com). LaTeX or Word are the preferred format, but PDFs are accepted if submitted with LaTeX code or a Word file of the text. There is no maximum limit, but recommended length is about 4,000 words. If needed, for editing considerations, the technical team may ask the author(s) to cut the article.
# Table of Contents

**Editorial**

*pag. 5*

### energy & commodity finance

**Security of Supply and the Cost of Storing Gas**  
Alvaro Cartea, James Cheeseman and Sebastian Jaimungal  
*pag. 7*

**Monetary Measurement of Risk: A Critical Overview - Part II: Coherent Measure of Risk**  
Lionel Lecesne and Andrea Roncoroni  
*pag. 13*

### banking & finance

**Towards a Theory of Internal Valuation and Transfer Pricing of Products in a Bank**  
Antonio Castagna  
*pag. 21*

### special interview

**Alessandro Mauro**  
LITASCO International Trading and Supply Company - Lukoil Group, Geneva  
*pag. 46*
Dear Readers,

we reached the third issue of Argo newsletter with great results. After the interesting contributions of the Winter publication, we are ready to reveal the latest developments in the fields of banking and energy finance.

This time we start with the Energy & Commodities finance section that raises with a contribution of three well-known personalities: Álvaro Cartea, James Cheeseman and Sebastian Jaimungal put forward a method devised to evaluate gas storage where a least-square Monte Carlo simulation method is used to solve the underlying optimization problem. It follows the second part of the crash course put forward by Lionel Lecesne and Andrea Roncoroni: these authors present the celebrated class of “coherent” measures of risk and provide the reader with a critical view of the axioms underlying this framework.

In this issue the Banking & Finance section is totally focused on transfer pricing topic. In an interesting contribution Antonio Castagna provides an innovative theory of internal valuation and transfer pricing of products in a bank.

In a dedicated section, Andrea Roncoroni interviews Alessandro Mauro (LITASCO company of Lukoil Group, Switzerland) about importance and evolution of financial and physical risk management in the energy sector.

We conclude by encouraging submission of contributions for the next issue of Argo Newsletter. Detailed information about the process is indicated at the beginning. Thanks for downloading Argo: feel free forwarding the publication’s web link to whoever might be concerned with any of the subject dealt with.

Enjoy your reading!

Antonio Castagna
Andrea Roncoroni
Luca Olivo
Energy & Commodity Finance

Energy Storage Valuation

Crash Course Part II: Monetary Measurement of Risk
The article offers an interesting insight on the importance of building gas storage facilities. There is a clear relationship between the capacity to store gas and the stability of gas prices. The focus of the research presented here is on the financial value of the gas storage: the authors show how to value storage facilities using Least Squares Monte Carlo (LSMC) without excluding realistic constraints from the valuation.

Álvaro CARTEA
James CHEESEMAN
Sebastian JAIMUNGAL

In 2009 Europe saw how the Ukraine-Russia gas crisis unraveled. After protracted negotiations over natural gas contracts, Russia’s energy company Gazprom cut gas supplies to the Ukraine which also affected large parts of southeast Europe. During this period, Ukraine had to survive on stored gas which would not last forever. These events highlighted the importance of guaranteeing stability of gas flows and that during times of duress, gas storage facilities play a decisive role to meet demand.

The case for building gas storage facilities is made on the back of both security of supply and to smooth prices at different time scales. Gas storage smooths prices across seasons because demand generally exhibits a strong seasonal pattern, and also smooths variations in prices across days by serving as a buffer to absorb unforeseen changes in daily supply and demand. Clearly, the greater the capacity to store gas, the more stable are prices.

However, there are two key questions. First, what is the cost of building a storage facility? Second, how much is a gas storage facility worth? The first question addresses the actual cost to build storage, and the second question, which is the subject matter of this article, addresses the financial value that is obtained from operating the facility. Therefore, only when the financial value of the storage is greater than the costs of building it, should storage be built.

In this article we show how to value a storage facility using Least Squares Monte Carlo (LSMC). We show how to incorporate realistic constraints in the valuation including: the maximum capacity of the storage, injection and withdrawal rates and costs, and market constraints such as bid-ask spread in the spot market and transaction costs.

Valuation of Gas Storage

Central to the problem of gas storage is its financial valuation not only for investment decisions, but also for operating and strategies, and hedging decisions for those wishing to trade around such assets. There are two approaches to the valuation of storage: intrinsic and extrinsic.

The value of a storage facility, like any option is its premium, which is made up of two parts, intrinsic and extrinsic. Relating this to a standard American option we have that the intrinsic value is the payoff from immediate exercise and the extrinsic value is the additional future exercise value of the option, the possibility of increasing the payoff.

The simplest approach to valuing storage, but one which ignores future exercise rights, is to compute the intrinsic value given the futures curve at the valuation date. This is the optimal combination of futures contracts which can be put in place, while respecting all the physical constraints of the storage.
facility, to extract value from the calendar spreads of the futures curve.

The extrinsic value represents the optionality in the physical asset and together with the intrinsic value this makes up the premium. The extrinsic value is that of the flexibility of the storage, the ability to trade the daily volatility. An extrinsic valuation can increase the value of storage from a few percent over the intrinsic to multiples of the intrinsic value depending upon the physical constraints of the facility and the dynamics of the underlying.

Least Squares Monte Carlo

Francis and Schwartz [5] originally applied the LSMC algorithm to valuing Bermudan options. Subsequently it has been extended and applied as a method for pricing swing options by Alfredo Ibáñez [1], and further to storage valuation by Carmona and Ludkovski [3] and Boogert and de Jong [2].

Reviewing the algorithm as presented in [5] we see that the intuition behind the least squares approach is best appreciated through the original application, an American-style option. The holder of such an option may exercise at any time prior to maturity, and at every potential exercise time the decision whether to exercise or not depends upon assessing the value of continuing without exercising, the continuation value, against that of exercising immediately, the intrinsic value. Longstaff and Schwartz use a simple least squares regression of the simulated future cashflows on to the current simulated stock prices to obtain this continuation value as a conditional expectation and hence the optimal exercise price (at that time) that maximises the option value.

As in any Monte Carlo simulation of a path dependent derivative the time to maturity is discretised, into k discrete potential exercise times

\[ 0 < t_1 \leq t_2 \leq \ldots \leq t_k = T \]

and the decision as to whether to exercise or not is taken at every point by comparing the intrinsic value with the value of continuation. Using the notation from [5] we have that when \( C(\omega; s; t, T) \) is the path of cashflows generated by the option, conditional on the option having not been exercised at or prior to time \( t \) and on the holder of the option following the optimal stopping strategy for all \( s \ (t < s \leq T) \), where \( \omega \) is a sample path, the value of continuation at time \( t_i \), can be expressed as the risk neutral expectation of the discounted future cashflows. We denote this continuation value as \( Cont(\omega; t_i) \) and the above argument is represented formally as

\[
Cont(\omega; t_i) = \mathbb{E}_Q \left[ \sum_{j=1}^{k} e^{-r(t_j-t_i)} C(\omega; s; t, T) \mid \mathcal{F}_{t_i} \right]
\]

where \( r \) is the risk free rate and \( Q \) the risk neutral pricing measure and \( \mathcal{F}_t \) is the filtration at time \( t \) representing the current futures forward curve.

LSMC works recursively through the discretised times using least squares regression to estimate the conditional expectation at \( t_{k-1}, t_{k-2}, \ldots, t_1 \). The LSMC is a recursive process as at each time the decision to exercise could change all the subsequent cashflows and therefore \( C(\omega; s; t_j, T) \) is not necessarily the same as \( C(\omega; s; t_{j-1}, T) \). The initial step is always to define the cashflows at the final timestep \( t_k \) and for an American option this is the intrinsic value of the option at maturity.

Longstaff and Schwartz give an example using the set of Laguerre polynomials as the basis functions for the regression and with these polynomials, the estimate of \( Cont(\omega; t_i) \) is then represented as:

\[
Cont_M(\omega; t_i) = \sum_{j=0}^{M} a_i L_j(X)
\]

where \( a_i \) and \( L_j \) are constant coefficients and the Laguerre polynomial terms respectively, \( M \) the number of basis functions and \( X \) the underlying spot prices. They report that if \( M > 2 \) the accuracy of the algorithm is not significantly dependent on the type or number of basis functions used, but more recent research show that accuracy does depend on the choice of basis functions, see [7].

In [5] \( Cont_M(\omega; t_i) \) is estimated by regressing the discounted values of \( C(\omega; s; t_j, T) \) onto the basis functions for the paths which are in the money, since the decision on whether to exercise or not is only required if the option has an intrinsic value greater than zero. Using only in-the-money paths reduces the number of basis functions required for an accurate estimation of the continuation values.\(^1\) With the conditional expectation estimated we can decide if exercise is optimal at \( t_i \) for each in-the-money path by comparing the intrinsic value against \( Cont_M(\omega; t_i) \). The algorithm continues recursively until we have the optimal decision for each point. We then discount the cashflow from each exercise point and take the average to find the value of the American put option, thus:

\[
C^A_t = \frac{1}{n} \sum_{p=0}^{p=n} e^{-r(t_p-t)} (K - S_{t_p}),
\]

where \( n \) is the number of paths.

\(^1\)In a storage problem we cannot make this simplifying assumption as all the cashflows which can be zero, positive or negative, must be considered.
It is important to note that when optimal exercise is indicated at $t_i$ the subsequent cashflows are altered to zero, as if exercise has taken place at $t_i$ then by definition it cannot take place at times $t_i < t \leq T$. It is these altered cashflows which feed into the next iteration where $\text{Cont}_M(\omega; t_{i-1})$ is estimated.

**Extending the LSMC to price Gas Storage**

Here we discuss how to use the LSMC approach in the valuation of gas storage, for other extensions and the pricing of gas interruptible contracts see [4]. If for example, the LSMC is used to price American-style options, the algorithm would require a matrix of two dimensions, time and number of simulated paths, where on each path an estimate of the optimal exercise time, a stopping rule for the option, is obtained. In application to multiple exercise problems we must create a third dimension, for swing options this is the number of exercise rights. We then proceed as follows, suppose we have a swing option with $N$ up-swing rights then we create a standard two dimensional LSMC matrix corresponding to each $N, N - 1, \ldots, 1$ rights remaining and working recursively, our initial step is to set the final cashflow conditions. These are different from the American option as we have multiple exercise opportunities and therefore if we have $N$ rights remaining the final $N$ timesteps in the $N^{th}$ matrix must be assumed to be exercise points, intuitively if the holder of a swing option is three days away from maturity and has three exercise rights left he will try and exercise on every day. The continuation values are estimated and then we must make the optimal exercise decision, but we must compare values from across the third dimension of our LSMC matrix.

Suppose we have $N$ swing rights remaining, if we want to decide whether to exercise one of our $N$ rights at time $t_i$, then we compare the value of not exercising and continuing with $N$ rights with the value of exercising and continuing with $N - 1$ rights. Introducing $\text{Cont}_M^N(\omega; t_i)$ as the continuation value with $N$ rights remaining at time $t_i$, and recall that $\omega$ is a sample path, we exercise a swing option with $N$ rights remaining at time $t_i$ if:

$$\text{Cont}_M^N(\omega; t_i) < I(\omega; t_i) + \text{Cont}_M^{N-1}(\omega; t_i)$$

where $I(\omega; t_i)$ is the intrinsic value of exercising. That is, if the value of the exercising one of our swing rights, $I(\omega; t_i)$ in addition to the value of continuing with one less right, $\text{Cont}_M^{N-1}(\omega; t_i)$ exceeds the value of not exercising and continuing with the current number of rights, then we exercise one right.

The many possible constraints of swing options, penalty functions, maximum number of upswings and downswings, all add complexity to the problem and change the way the algorithm proceeds. For instance, if a penalty is payable when the maximum number of swings is not reached then the final cashflow may be negative as it may be optimal to pay the penalty rather than suffer a loss from exercising, for other approaches see [6].

**Storage LSMC**

A storage facility is much like a swing option with an up- and down-swing option for every day of the contract, for instance, on any day we may buy gas and inject it into storage or withdraw gas and sell it to the market. The injection and withdrawal limits are governed by ‘ratchets’ which are volume dependent and unique to each type of storage facility. These occur as the inventory level increases and it becomes harder to inject more gas and as the inventory decreases it becomes harder to withdraw gas.

Our action is also governed by minimum and maximum inventory levels that apply across the duration of the storage contract, and initial and final conditions which stipulate the inventory state in which the facility is obtained and must be returned. There may also be a penalty, dependent on the volume level and spot price for breaking any of these constraints. The type of facility may also dictate fixed injection and withdrawal costs which pay for the additional energy required to operate in injection or withdrawal mode. In addition to the physical constraints we are also governed by market constraints such as a bid-ask spread, transaction costs and of course the discount rate.

To value a storage facility we again set up a three dimensional LSMC matrix with the third dimension being the volume of gas in the storage facility. In this way we can accommodate the above constraints by making the appropriate comparisons with continuation values across the different volume levels.

So working recursively from maturity back to time zero, at inventory level $V$, we compare three possible actions, inject volume $v_i$ and continue at inventory level $V + v_i$, take no action and continue at inventory level $V$, or withdraw volume $v_w$ and continue at inventory level $V - v_w$. Obviously the physical constraints above, ratchets, max and min inventory levels and the initial and terminal constraints define the allowed sets of $v_i$ and $v_w$ from which we can choose.
To summarise we value a storage facility subject to the following constraints. $V_{\text{max}}$: The maximum capacity of the storage facility. $V_i$: The inventory level at which the storage contract is initiated. $V_f$: The inventory level at which the storage contract must be returned at the end of the contract. $v_{\text{inj}}(V)$ & $v_{\text{withdraw}}(V)$: The ratchets, i.e. the inventory level dependent injection and withdrawal rates. $C_{\text{inj}}$ & $C_{\text{withdraw}}$: Fixed injection and withdrawal costs in pence per therm. $BA$, $C_{\text{trans}}$ & $r$: The market constraints, bid-ask spread, transaction costs and the discount rate respectively.

**Storage Value**

We value a contract on a storage capacity of 1 million therms with daily injection and withdrawal rates of 100,000 therms, we receive and must return the facility empty. We value a contract duration of 1 year, with a daily trading frequency and zero interest rates, across a range of contract start dates from 1st January to June 30th 2008. In figure 1 we show the intrinsic value and the premium with and without jumps added to the price process, we also show the futures curve to which the spot price was calibrated.

In Figure 1 we see that the most basic indicator of the storage value, the intrinsic value, is constant across the contract start dates. The intrinsic value is the optimal combination of futures contracts given a static view of the futures curve on a particular, and subject to the physical constraints of the storage facility. As the gas futures curve is dominated by a repeating seasonal pattern the relative difference in prices across the year is constant as we change the start date of our storage contract. This results in a constant intrinsic value regardless of any contango or backwardation inherent in the market.

In the range of Spring start dates, where we see a higher spot price, the premia (with and without jumps) are lower than spring and summer start dates. When we start a storage contract with zero inventory our first action can only be to inject gas. If we start the contract in Spring we wait until the low summer prices to fully fill our facility then selling the following Spring. If, in contrast, we start a contract in the summer the situation is reversed and we immediately fill our storage and sell the following Spring. Throughout the duration of both these cases we trade not only the overall calendar spread but the daily volatility as well. The calendar spread is constant regardless of when the contract starts as it depends only the relative difference of summer and Spring prices. This implies that the difference in storage premia across the year must arise from the way in which the daily volatility is traded, specifically from the increased flexibility of having more gas at an earlier date in the case of a summer start contract.

In respect of the jumps our path simulation with jumps give spot prices in the range of 150-200 pence per therm, and approximately 7 jumps per year. This is in comparison to normal prices in the range of 7-100 pence per therm. With this comparison in mind we see that the majority of the extrinsic value comes from the volatility of the price process rather than the jumps, as the extrinsic value shows an increase of $2.5 \times 10^7$ over the intrinsic value, whereas the extrinsic with jumps shows only...
a further $1 \times 10^7$ increase.

ABOUT THE AUTHORS
Álvaro Cartea University College London
Email address: a.cartea@ucl.ac.uk

James Cheeseman BP plc
Email address: jimcheeseman@hotmail.co.uk

Sebastian Jaimungal University of Toronto
Email address: sebastian.jaimungal@utoronto.ca

ABOUT THE ARTICLE
Submitted: January 2014.
Accepted: February 2014.

REFERENCES


The research and consulting network in energy and commodity finance

People
A team of top academics and practitioners working on quantitative modelling, energy, and commodity risk.
Know more about us >

Articles
A continuous flow of papers concerning, developing, and evaluating cutting-edge techniques in commodity finance.
Follow our publications >

ArGo Newsletter
The digital quarterly magazine on best practices and models for bank and corporate risk management.
Check latest issues >

Books
Comprehensive monographs encompassing key topics in physical and financial risk management.
See more about our volumes >

Garp ERP®
Full training in preparation to the Garp Energy Risk Professional (ERP®) certification exam.
See GARP ERP training details >

Programs
Academic programs offering high level knowledge and skills in energy and commodity finance.
Check academic programs list >

Codes
A basket of computer codes for effectively pricing, hedging, and managing energy and commodity-based positions.
Download computer codes >

FloRisk™
A web-based system of advanced solutions for pricing, monitoring, and managing energy and commodity risk.
Get more about our solution >

The New Frontier of Cloud Consulting

Free registration

© All rights reserved Follow us on www.energisk.org or contact us at contact@energisk.org
Monetary Measurement of Risk: a Critical Overview

Part II: Coherent Risk Measures

In this second part of the course Lionel Lecesne and Andrea Roncoroni introduce the subclass of coherent measures of risk introduced by Artzner et al. (1999), providing with specific properties, definitions and examples from this class.

Lionel LECESNE
Andrea RONCORONI

In Lecesne and Roncoroni [4], we introduce the notion of monetary measure of risk borne by any financial claim. Our presentation moves from general definitions to concrete instances, including the benchmark measure Value-at-Risk (VaR). Part II develops a treatment of the class of coherent (monetary) measures of risk put forward by Artzner et al. [2]. Our goal is to illustrate the main features of this class of measures in light of application to practical cases. In this respect, we put our focus on the ambiguity of the term “coherent” and show that the connotation of consistency it naturally embeds is more an issue of lexical interpretation than actual meaning. As an example, we show that lack of subadditivity of VaR (which prevents from it to being a “coherent” measure of risk) is more a desirable property than a drawback, as is claimed in most of existing sources in the specialized literature.

Coherent Measures of Risk

At the end of the nineties, risk measurement theory was suffering from lack of axiomatic foundations. Artzner et al. [2] propose a number of “economically desirable” axioms that a “coherent” monetary measure of risk $\rho$ ought to satisfy. These properties lead to the definition of a class of risk measures referred to as coherent by the authors.

Since “coherent” measures of risk constitute a subclass of monetary risk measures, they obviously meet the anti-monotonicity and translational anti-variance requirements. We reproduce their statement here for the reader’s convenience (Lecesne and Roncoroni [4]):

- **Anti-monotonicity (AM)**: for any pair of positions $X, Y \in \mathcal{X}$, with $X \leq Y$:
  $$\rho(X) \geq \rho(Y).$$

- **Translational anti-variance (TA)**: for any pay-off $X \in \mathcal{X}$ and cash amount $m > 0$:
  $$\rho(X + m) = \rho(X) - m.$$

To these properties, one requires that combining positions does not generate additional risk. Embrechts et al. [3] provide three qualitative arguments underpinning this assumption, which is often labelled as “subadditivity”:

1. **Diversification effect.** The property reflects a well-known principle whereby risk can be reduced through diversification;

2. **Merging incentive.** If a regulator imposes a measure of risk which might entail risk increase upon merging, then financial institutions might get an incentive to break up into subsidiaries in order to reduce regulatory capital requirements stemming from risk assessment.
3. RM decentralization. Subadditivity ensures that the risk borne by any financial institution is superiorly bounded by the sum of risk figures of composing desks. Therefore, limiting these latter offers control to the overall risk, thus making of risk decentralization an effective risk management policy.

If we denote a coherent measure of risk as $\rho^{coh}: \mathcal{X} \to \mathbb{R}$, the previous property can be stated as follows:

**Subadditivity (SUB):** for any pair of positions $X, Y \in \mathcal{X}$,  
$$\rho^{coh}(X + Y) \leq \rho^{coh}(X) + \rho^{coh}(Y);$$

One further asks for factoring the size of a position out of the corresponding risk assessment. In particular, doubling a financial position exactly doubles the initial risk.

**Positive homogeneity (PH):** for any $X \in \mathcal{X}, \lambda \geq 0$,  
$$\rho^{coh}(\lambda X) = \lambda \rho^{coh}(X).$$

We are led to the following:

**Definition.** A Coherent Measure of Risk $\rho^{coh}: \mathcal{X} \to \mathbb{R}$ is a subadditive and positively homogenous monetary measure of risk.

**Spectral Measures of Risk**

**Expected Shortfall (or, Conditional VaR)**

Expected Shortfall (ES) is probably the most used risk measure after VaR. It is anti-monotone, translation-antivariant, and positively homogenous. The main theoretical advantage over VaR is that it is subadditive, hence coherent. From a practical point of view, a major benefit over VaR lies in the kind of information it provides to the user. Recall that VaR delivers an assessment about the minimal size of losses a portfolio may experience at a defined level of confidence. Consequently, no information is provided about the severity of the actual losses exceeding a VaR threshold. Given a confidence level, ES provides a first estimate of this severity by delivering the expected value of losses conditional to exceeding the VaR threshold.

VaR may be defined either as the $\alpha$-quantile of a loss distribution, i.e., $F_X^{-1}(\alpha) := \inf\{x \in \mathbb{R} : P[X \leq x] \geq \alpha\}$ or as the $(1 - \alpha)$-quantile of the P&L distribution, i.e., $-F_X^{-1}(1 - \alpha) := -\inf\{x \in \mathbb{R} : P[X \leq x] \geq 1 - \alpha\}$, where $\alpha \in [0, 1]$ is generally taken to be close to 1. In what follows, we will retain the second of these definitions.

**Definition.** For a random P&L $X \in \mathcal{X}$ with continuous distribution function and a confidence level $\alpha \in [0, 1]$, the 100$\alpha$% Expected Shortfall of $X$ is defined as the expected loss (i.e., minus P&L) conditional to it exceeding the 100$\alpha$% Value-at-Risk:

$$ES(X) := -E\{X|\{X \leq -VaR_\alpha(X)\}\},$$  \hspace{1cm} (1)

where $VaR_\alpha(X) = -F_X^{-1}(1 - \alpha)$. Equation (1) clearly shows the role played by VaR in the definition of ES, which is sometimes referred to as Conditional Tail Expectation or Conditional VaR. A more general, although less immediate to grasp, definition of ES is the following one:

$$ES_\alpha(X) := -\frac{1}{1 - \alpha} \int_0^1 F_X^{-1}(1 - u)du$$
$$= \frac{1}{1 - \alpha} \int_0^1 VaR_u(X)du,$$  \hspace{1cm} (2)

where the link to VaR is ever more evident. This second definition is less restrictive since it does not require the distribution function to be continuous.

Next, we provide an example based on a simple position showing the way ES varies across confidence level and volatility of the underlying asset.

**An Illustrative Example**

A trader buys 1,000 stock shares at a spot price $S_0 = $100. He wishes to assess the regulatory capital associated to his $100,000 investment over a management horizon of $T$ time units. Estimations based on past data indicate that absolute price changes $\Delta S := S_T - S_0$ follow a Normal law with zero mean and variance given by $\sigma^2$. Position’s P&L’s $X := 1000\Delta S$ are Normal as well, with zero mean and variance equal to $(1000\sigma)^2$. If capital requirement is determined by ES, then formula (2) reads as:

$$ES_{\alpha,\sigma}(X) = -\frac{1}{1 - \alpha} \int_0^1 1000F_{\Delta S}^{-1}(1 - u)du,$$
$$= -\frac{1000\alpha}{1 - \alpha} \int_0^1 \Phi^{-1}(1 - u)du,$$

where $F_{\Delta S}$ is the cumulative probability distribution of $\Delta S$, $\Phi$ denotes the standard normal distribution, and the subscripts explicitly indicate dependence on confidence level $\alpha$ and volatility $\sigma$.

Figure 1 exhibits a 3-dimensional plot of ES across varying levels of confidence threshold and asset volatility. We note that:
FIGURE 1: ES of a normal centered P&L with varying volatility (sigma) and confidence level (alpha)

FIGURE 2: VaR, ES and SMRs of a normal centered and reduced P&L
1. ES is positive, and so is capital requirement. In other terms, cash withdrawal is never authorized, as opposed to what might be the case if one uses VaR assessments.

2. ES increases along with confidence level \( \alpha \). For a standard deviation \( \sigma = 1 \), we have \( \text{ES}_X(\alpha = 0.98, \sigma = 1) = $2,401.1 \) and \( \text{ES}_X(\alpha = 0.99, \sigma = 1) = $2,625.63 \).

3. ES is proportional to asset volatility. For a confidence level \( \alpha = 0.95 \), an increase of \( \sigma \) from 1 to 5 entails an economic capital rise from $2,054.8 to $10,274.

General Spectral Measures of Risk

Acerbi [1] generalizes the notion of ES and puts forward the class of Spectral Measures of Risk (SMR). The idea is to compute an appropriate weighted average of all possible payoffs stemming from a given position. Weighing is performed in a way to emphasize worst outcomes over best ones. The resulting figure turns out to be “coherent” in the sense of Artzner et al. [2].

**Definition.** For a random P&L \( X \in \mathcal{X} \) and generalized inverse cumulative distribution function \( F_X^{-1}(u) := \inf \{ x \in \mathbb{R} : F_X(x) \geq u, u \in [0, 1] \} \), a **Spectral Measure of Risk** of \( X \) is defined as:

\[
\text{SMR}_w(X) := -\int_0^1 w(u) F_X^{-1}(u) du \quad (3)
\]

for each positive weighing function \( w : [0, 1] \rightarrow \mathbb{R} \) satisfying:

1. \( \int_0^1 w(u) du = 1 \) and
2. \( w(u_1) \geq w(u_2) \) for any \( 0 < u_1 < u_2 \leq 1 \).

**Remark.** The weighing function \( w \) is called “risk spectrum” and properties 1 and 2 ensure that SMR are “coherent” (Acerbi [1]).

**Example.** A flexible class of risk spectra is given by the parametric set of functions:

\[
w_\gamma(u) := \frac{\exp(-u/\gamma)}{\gamma(1 - \exp(-1/\gamma))}, \quad (4)
\]

where \( \gamma > 0 \) is the parameter allowing for tuning weight assignment to severe losses (the lower \( \gamma \), the stronger the weight).

**Example.** Risk spectrum \( w_\alpha(u) = (1/1 - \alpha)1_{u \leq 1 - \alpha} \) delivers ES:

\[
\text{SMR}_{w_\alpha}(X) = -\frac{1}{1 - \alpha} \int_0^{1-\alpha} F_X^{-1}(u) du =: \text{ES}_\alpha(X).
\]

While ES averages losses beyond VaR threshold, a general SMR scans the whole set of quantiles. A key practical feature of SMR is that the user (or regulator) may tune the weight assignment across the set of potential losses to reflect his own aversion to risk.

A Numerical Example

We now put forward a pictorial comparison of VaR, ES, and some other SMR’s. Figure 2 exhibits figures corresponding to VaR and ES across varying levels \( \alpha \) of confidence over the interval \([0, 1]\). These graphs are then superimposed to those related to two SMR’s. These latter stem from adopting weighing functions defined through formula (4) with tail parameter \( \gamma = 0.1 \) and \( \gamma = 1 \). These parametric instances correspond to great and, respectively, low importance assigned to high level of potential losses.

A few comments of this picture are worth of noticing:

- ES always exceeds VaR. This property is rather obvious. The key point is the main consequence it entails: ES is more conservative than VaR, in that it always provides a higher capital requirement to the user.

- VaR is negative for some values of the confidence level \( \alpha \) (here \( \alpha \leq 0.5 \)), whereas ES is always positive. Again, this observation reflects the idea that ES is more prudent than VaR. Indeed, assuming that the portfolio manager would choose a such low confidence level, the negative outcome of VaR would allow him to withdraw cash from the position (and hence increase the risk due to the cash translational anti-variance property) until VaR attains value zero. On the contrary, ES is never negative which means that it does not authorize to withdraw cash from the position.

- The exact location of SMR relative to both VaR and ES depends on the value of the tail parameter \( \gamma \). As previously remarked, regulatory capital stemming from SMR increases as long as stronger weight is assigned to high losses: this corresponds to values of the tail parameter \( \gamma \) near 0.
On the Consistency of Risk Coherence

VaR may be superadditive. Hence it is not a "coherent" measure of risk, in the sense of Artzner et al. [2]. An example helps at clarifying why risk superadditiveness may well be acceptable.

Let banks A and B dispose of $10,000 available for lending. Assume they follow distinct credit lending policies: bank A lends the total amount to a single client, whereas bank B grants $1,000 to each of ten individual clients. Let all debtors be identical, exhibit a probability of default equal to 0.02, and assume one’s default has no impact onto the default probability of the others. The paid back sum amounts to the borrowed cash plus a 5%-interest payment. For reader’s ease of comparison, we reproduce here the definition of VaR:

\[
\text{VaR}_\alpha(X) := \inf \{ m \in \mathbb{R} : P[-X \leq m] \geq \alpha \}. \quad (5)
\]

Let \( X_A \) denote the bank A random P&L resulting from lending. If the sole client defaults, they lose $10,000; otherwise, a benefit of $10,000×5 % = $500 occurs upon repayment. We have:

\[
\begin{align*}
\mathbb{P}[X_A = -500] &= 0.98 \quad \text{(no default)} \\
\mathbb{P}[X_A = 10,000] &= 0.02 \quad \text{(default)}
\end{align*}
\]

and nesting (6) into formula (5) leads to \( \text{VaR}_{0.98}(X_A) = -$500 \) and \( \text{VaR}_{0.99}(X_A) = $10,000 \).

Let us now look at the risk borne by bank B. Let \( X_i \) denote the bank B random P&L stemming from debtor \( i = 1, \ldots, 10 \). If individual \( i \) defaults, the bank loses the whole outstanding capital, hence \( X_i = -$1,000 \). If they honor their debt, then the bank’s profit equals interest and \( X_i = $50 \). Let \( D_i \) be a random variable assuming value 1 in case individual \( i \) defaults and 0 otherwise. Equation (7) then synthesizes bank B’s P&L relative to the debt of client \( i \):

\[
X_i = 50 \times (1 - D_i) - 1,000 \times D_i. \quad (7)
\]

Since bank B has 10 clients, their total P&L equals \( X_B = \sum_{i=1}^{10} X_i \). In regard to equation (7), we may rewrite:

\[
X_B = 500 - 10,500 \sum_{i=1}^{10} D_i. \quad (8)
\]

where \( \sum_{i=1}^{10} D_i \) follows a binomial law with a number of trials equal to 10 and probability of success (i.e., default) equals 0.02\(^2\). Recall that formula (5) is based on the loss distribution, we hence consider

\[
-X_B = 10,500 \sum_{i=1}^{10} D_i - 500. \quad (9)
\]

\( \text{VaR}_\alpha(\sum_{i=1}^{10} D_i) \) corresponds to the \( \alpha \)-quantile of the cumulative distribution of \( \sum_{i=1}^{10} D_i \). From a binomial probability table, we get \( \text{VaR}_{0.98}(\sum_{i=1}^{10} D_i) = 1 \) and \( \text{VaR}_{0.99}(\sum_{i=1}^{10} D_i) = 2 \). Finally, by replacing these values into (9), we find that \( \text{VaR}_{0.98}(X_B) = $10,000 \) and \( \text{VaR}_{0.99}(X_B) = $20,500 \).

In agreement with the idea that diversification reduces risk, bank B has, unlike bank A, cut the $10,000 in order to grant them to several clients. Paradoxically, VaR assessment yields greater risk to bank B than to bank A. This is a consequence of non-subadditivity of VaR. Table 1 reports VaR outcomes for this example.

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>( \alpha = 0.98 )</th>
<th>( \alpha = 0.99 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{VaR}_\alpha(X_A) )</td>
<td>-500</td>
<td>10,000</td>
</tr>
<tr>
<td>( \text{VaR}_\alpha(X_B) )</td>
<td>10,000</td>
<td>20,500</td>
</tr>
</tbody>
</table>

**Table 1**: VaR outputs for banks A and B with varying confidence level.

For both values of confidence level \( \alpha \), capital requirement for bank B is more than twice the one for bank A. It is striking to observe that bank A negative VaR corresponding to \( \alpha = 0.98 \) implicitly allows an increase of risk exposure by $500, in spite of the fact that bank B is required to reduce exposure by $10,000.

According to Artzner et al. [2], financial risk metrics ought to be subadditive, among others. This property is grounded on the intuitive assumption whereby gathering two positions in a single portfolio would eventually reduce risk through a diversification effect. In our view, this property should not be universally accepted: although combination often reduces overall risk, it may increase it as well. Examples ranging from corporate finance to energy economics and liquidity theory show that a ‘snowball’ effect resulting from combining positions may lead to an increase of expected losses and the corresponding capital requirements or reserves. The dependence structure between financial positions ultimately constitutes the reason why subadditivity (and positive homogeneity) may need fail upon risk assessment. Liquidity related issue are another source of risk superadditivity: doubling the size of an illiquid portfolio more than doubles the risk from holding the portfolio. An

\(^2\)Formally, we denote it as: \( \sum_{i=1}^{10} D_i \sim B(10;0.02) \).
example may help at clarifying the latter point.

Consider an investor who needs to immediately cash a portfolio of identical bond issues. Assume that the buy side of the bond order book is described by the following bids: the best bid quotes $1,000 for each of the first 50 contracts; the second bid quotes $990 for each of the following 100 contracts; the third bid quotes $970 for each of the last 200 contracts. Assume that the risk borne by the portfolio holder is represented by the difference between the value of the portfolio quoted at the best bid and the effective liquidation value. Let the investor hold 100 bond issues. Given the previous description, the fictitious value of this portfolio at best bid is equal to $100,000. Its liquidation value is equal to $99,500. Hence, they would incur into a lost of $500 due to a liquidity issue. Now assume that the standing portfolio doubles to 200 bonds. The value at best bid is now $200,000, whereas the total liquidation value equals $197,500. Now the cost of liquidity is $2,500. We see that positive homogeneity fails. Moreover, risk assessment moves from $500 to $2,500, which exceed twice the initial figure, that is $1,000. Hence, subadditivity fails too. This example shows liquidity issues may reasonably require a superadditive risk assessment in other to appropriately represent the exposure of combined positions.

We may then conclude with a statement of a consistency principle: subadditivity and superadditivity should be properties of the underlying positions and their dependence structure as opposed to being taken as an axiom fulfilled by the adopted risk metrics. As long as VaR does not impose any condition about either subadditivity or superadditivity, we believe it complies with the mentioned consistency principle in sharp contrast to what “coherent” measures of risk do.

ABOUT THE AUTHORS
Andrea Roncoroni is full professor of Finance at ESSEC Business School, Paris, and Visiting Fellow at Bocconi University, Milan. Research interests are related to energy and commodity finance, risk monitoring and mitigation and applied quantitative modelling.
Email address: roncoroni@essec.edu

Lionel Lecesne is Ph.D. candidate at University of Cergy, France. He was formerly student at Ecole Normale Supérieure in Cachan, France. Research interests include risk measurement, energy finance and financial econometrics.
Email address: lionel.lecesne@u-cergy.fr

ABOUT THE ARTICLE
Submitted: March 2013.
Accepted: April 2014.

References


Advisory Research and Training in Finance

**Concentric** is a company specialized in the advisory, research and training in sales and finance. The company adopts a dynamic approach to its organisational evolution, reflecting ongoing changes in the business environment, the requirements of clients and developments in best practices.

**QUANTITATIVE ADVISORY**

**quad** is the fusion of Concentric and Iason international risk advisory, research and training practitioner teams, rigorous project methodologies and tried-and-tested tools.

**CORPORATE GOVERNANCE ADVISORY**

**cga** comprises organisational analysis, design and development projects for bank, insurance company and corporate compliancy.

**CONCENTRIC RESEARCH**

**core** includes service consultation, needs analysis, project modeling, satisfaction analysis, industry survey and financial analysis.

**FINANCE MASTER CLASS**

**fmc** is one of Europe’s most appreciated and successful public technical financial training programs for banking and insurance professionals.

**SALES AND FINANCE EDUCATION**

the **safe** team adopts a holistic and value-adding approach to the design and development of in-company finance, commercial and management training projects.

For more information visit www.concentric-italy.com
Banking & Finance

Transfer Pricing
Towards a Theory of Internal Valuation and Transfer Pricing of Products in a Bank

Funding, Credit Risk and Economic Capital

In this article Castagna outlines a framework to theoretically identify the components of the value that a bank should attach to a deal. Moreover the framework suggests how to charge these components to the relevant departments and/or to the final counterparty (client) by an internal transfer pricing system.

Antonio CASTAGNA

The pricing of contracts in the financial industry relies on theoretical results in many cases. Suffice to remember that the valuation of derivative contracts hinges on the results of Option Pricing Theory, mainly developed between the 1970’s and the 1990’s.

In some cases the practice differs from the theory. For example, the inclusion of funding costs in the valuation of loans has been seen in striking contrast with the results proved by academicians such as Modigliani and Miller, awarded with Nobel prized for their achievements.

In the last few years, the debate has sparked also as far as derivative contracts are concerned: there is not a widespread agreement on whether adjustments such as the Debit Value Adjustment (DVA) and Funding Value Adjustments (FVA) are justified or not. Some authors have proposed full pricing approaches including every possible adjustment in the value: we can refer to the most recent works in this area by Brigo, Morini and Pallavicini [1] and Crepy et al. [11]. It should be noted that these authors are more interested in finding out a valuation formula encompassing all the components of the pricing as seen by both counterparties, but they do not investigate which components can be effectively replicated so that they represent a recoverable production cost. In few words, these approaches aim at determining a fair tradable price on which both parties may agree by acknowledging each others’ risks and costs, but they cannot always be correct and consistent methods to evaluate derivative contract once they are traded and entered in the bank’s books.

The heart of the matter with the revaluation of derivative contracts is in the assumptions the bank makes about its hedging policies: they can be more or less realistic and feasible, so that different evaluations can be obtained which have to be challenged by the ability to actually implement the assumed hedging strategies. In this work we will try to shed some lights on how products (contracts) should be internally evaluated by a bank, which are the costs that should be transferred to the counterparty, the extent to which the theoretical results still applies in practice and under which conditions they do so. To this end, we will refer to a stylised balance sheet in a basic multi-period setting.

Some of the results we will crop up in the present work are derived in Castagna [9], but within a framework that, although similar to the one we will sketch below, was not capable to avoid an aporetic situation that was however acknowledged in the final part of that paper. By modifying and expanding said framework, we will hopefully provide
a firmer theoretical ground to some of the claims made elsewhere in Castagna [6] and confirm the other results in Castagna [7].

Balance Sheet with a Single Asset

Let us start with the assumption that we are in an economy with interest rates set constant at zero level. An asset $A_1(t)$ has an initial price $X_1$ at time $t$, terminal pay-off $A_1(T) = X_1 \times (1 + s_1)$. For simplicity, we will assume that $T - t = 1$ in what follows. There is a probability $PD_1$ that the asset’s issuer defaults between times $t$ and $T$, in which case the buyer of the asset will receive in $T$ a stochastic recovery $Rec_i$ expressed as a percentage of the face value $X_1$. $Rec_i$ can take values $Rec_1$, for $j = 1, ..., J$, and each possible value can occur with probability $p_j$.

$$A_1(t) = E[X_1 \times (1 + s_1)]$$
$$= X_1 \times (1 + s_1)(1 - PD_1) + X_1Rec_1PD_1 = X_1 \quad (1)$$

where $Rec_1 = \sum_{j=1}^{J} Rec_j p_j$ is the expected recovery in the event of the issuer’s bankruptcy.

Assume that an investor (with zero leverage) buys the asset $A_1$. In perfect markets, the spread $s_1$ they require from the issuer of the asset $A_1$ is simply the fair credit spread $cs_1$ remunerating the credit risk:

$$s_1 = cs_1 = \frac{Lgd_1PD_1}{1 - PD_1} \quad (2)$$

where $Lgd_1 = (1 - Rec_1)$ is the loss given default rate (the complementary to 1 of the recovery rate). The spread $s_1$ is set at the level that makes the terminal expected value of $A_1$ equal to the present value, $X_1$.

Assume now that a bank invests in the same asset $A_1$. The bank uses leverage, which means that it issues a bond to buy the asset. We denote the value of the debt at time $t$ with $D_1(t)$ and of the equity with $E$. The equity is invested in a risk-free bank account, $B(t) = E$. The amount needed to fund the asset is $D_1(t) = X_1$; this amount is raised by the bank with a bond issuance. The terminal pay-off of the bank’s debt is linked to the pay-off of asset $A_1$. In fact, the debt $D_1$ pays at maturity $X_1 \times (1 + f_1)$ if the asset’s issuer does not go defaulted, or $X_1Rec_1 + E$ (for each possible recovery rate) if it does. In a perfect market, with perfectly informed agents, the fair bank’s funding spread, requested by the buyers of the bank’s bond, is derived from the following equation:

$$D_1(t) = E[X_1 \times (1 + f_1)]$$
$$= X_1 \times (1 + f_1)(1 - PD_1) + \sum_{j=1}^{J} \min[X_1Rec_j, E]X_1 \times (1 + f_1)]p_jPD_1$$
$$= X_1 \quad (3)$$

The LHS of equation (3) is the expected value of the bank’s bond at expiry: if the issuer of the asset $A_1$ survives, then the bank is able to reimburse the notionial of the bond plus the funding spread $X_1 \times (1 + f_1)$. If the assets’ issuer goes bust, then the bond will be reimbursed with the funding spread only if this amount is smaller than the recovery value of the asset plus the initial equity deposited in the risk-free account: this is computed for each possible value of $Rec_1$, and weighted for the corresponding probability. This expected value must equal the present value of the debt, which is the face amount $X_1$.

We assume that there is at least one case when $E < X_1Lgd_1 + X_1f_1$ (the equity is not able to fully cover the loss given the default of the asset’s issuer and the funding cost). The funding spread is the level of $f_1$ equating the expected value at expiry with the present value:

$$f_1 = \frac{Lgd_1PD_1}{1 - PD_1} \quad (4)$$

where $Lgd_1 = (1 - \sum_{j=1}^{J} \min[X_1Rec_j, E]X_1 \times (1 + f_1)]/X_1$.

3The quantity $Lgd_1$ can be also be seen as the expected loss given default on the bank’s debt, $Lgd_1 = Lgd_{D_1}$, and the corresponding recovery is $Rec_{D_1} = \sum_{j=1}^{J} \min[X_1Rec_j, E]X_1 \times (1 + f_1)]/X_1$. They are different from the loss given default and the recovery of the asset $A_1$ since when the asset’s issuer defaults, a fraction of the loss is covered by the equity $E$, which is also the maximum amount the shareholders of the bank are liable for. Since $Lgd_1 > Lgd_1 = Lgd_{D_1}$, then $s_1 > f_1$, the bank “enjoys” (if an economic agent may ever enjoy from losing a part of its own capital) from the fact that the equity is contributing to the coverage of the potential future credit losses. If $E > X_1Lgd_1 + X_1f_1$ for all possible $Lgd_1$, then $Lgd_1 = Lgd_{D_1} = 0$ and $f_1 = 0$.

Assume now that the bank has a bargaining power in setting the fair rate $s_1$, so that it is able to set it at a level different from the fair level $cs_1$ required by a non-leveraged investor. In this way the
bank tries to cover costs and losses other than the credit losses. The fair mark-up spread $ms_1$ the bank has to charge on asset $A_1$, given the limited liability of the shareholders, is obtained by the equation:

$$
\text{VB}(t) = E[X_t \times (1 + ms_1) + E - X_t \times (1 + f_1)] = [(ms_1 - f_1)X_t + E](1 - PD_1) + \sum_{j=1}^{J} \max[X_t \text{Rec}_j + E - X_t \times (1 + f_1) ; 0]p_1 PD_1 = E
$$

Equation (5) is the net value of the bank at time $t$, which is equal to the expected value of the future value in $T$ of the bank’s total assets (the final asset’s pay-off $X_t$ plus the margin $ms_1X_t$, plus the equity amount deposited in the risk-free account $E$), minus the bank’s total liabilities (the notional of the debt $X_t$ plus the funding costs $f_1X_t$). Since the bank’s shareholders invested the initial amount $E$, the expected net value $\text{VB}(t)$ must equal $E$ to prevent any arbitrage.

Assume $E < X_tLgd_{j_1} + X_tf_1$ for one or more $Lgd_{j_1}$’s. Indicating with

$$
\mathcal{R}_1 = \sum_{j=1}^{J} \max[X_t \text{Rec}_j + E - X_t \times (1 + f_1) ; 0]p_1,
$$

the mark-up spread, from (5), is:

$$
ms_1 = \frac{E - \mathcal{R}_1}{1 - PD_1} + f_1 = cs_1^+ + f_1
$$

This is the sum two components:

- the “adjusted” credit spread $cs_1^+ < cs_1$ on the asset $A_1$, due the loss given default $(E - \mathcal{R}_1)/X_t$ lower than $Lgd_{j_1}$. The smaller loss given default suffered by the shareholders is generated by the leveraged investment in the asset $A_1$, and by the limited liability up to the equity amount $E$. In practice a share of the losses given default is taken by the debt holders, whence the lower credit spread;

- the funding spread $f_1$ paid by the bank on its debt.

By some manipulations, it is easy to check that in perfect markets where the credit spreads set by investors are fair and given in (2), we have:

$$
ms_1 = \frac{E - \mathcal{R}_1}{1 - PD_1} + \frac{Lgd_{j_1}PD_1}{1 - PD_1} = \frac{Lgd_{j_1}PD_1}{1 - PD_1} = s_1
$$

So the mark-up spread is just the credit spread of the asset $A_1$ required by a for a non-leveraged investor.

If $E > X_tLgd_{j_1} + X_tf_1$ for all possible $Lgd_{j_1}$ (the equity is large enough to cover the loss given the default of the asset’s issuer and the funding costs in each possible case), then $f_1 = 0$ and $\sum_{j=1}^{J} \max[X_t \text{Rec}_j + E - X_t \times (1 + f_1) ; 0]p_1 = X_t\text{Rec}_1 + E - X_t \times (1 + f_1)$. From equation (5) we have that the mark-up spread is $ms_1 = \frac{(Lgd_{j_1}PD_1)}{(1 - PD_1)}$, which is consistent with the stated results.

**Proposition 1.** If the bank holds only one asset, the leverage is immaterial in its internal pricing. Differently stated, the bank can price the asset as it were an non-leveraged investor, and the assets’ price would depend only on its expected future pay-off.

The bank’s default probability $PD_1$ is linked to the issuer’s probability of default $PD_{j_1}$, and it depends on the amount of equity that can be used to lower the $Lgd_{j_1}$ in the different cases. In general terms, the bank defaults when its value is below zero: since the shareholders’ liability is limited, the bank’s value is floored at zero, so that the bank goes bust when $R_1 = \max[X_t \text{Rec}_j + E - X \times (1 + f_1) ; 0] = 0$, for one or more $j$’s. We have that:

$$
PD_B = \sum_{j=1}^{J} 1_{\{R_1 = 0\}}p_1 PD_1
$$

where $1_{\{\theta\}}$ is the indicator function.

The conclusion is that bank’s leverage and funding costs are immaterial in pricing the asset $A_1$: the requested spread to make fair the contract is the same both for a non-leveraged investor and for the leveraged bank. This result is definitely not new: in a different setting, we reached the same conclusions as the well known works by Modigliani&Miller (M&M) [18] and Merton [16], whose results remain fully valid also in our setting.

When we are not in perfect markets, formula (6) provides the bank with a useful tool to price the asset $A_1$: the requested spread to make fair the contract is the same both for a non-leveraged investor and for the leveraged bank. This result is definitely not new: in a different setting, we reached the same conclusions as the well known works by Modigliani&Miller (M&M) [18] and Merton [16], whose results remain fully valid also in our setting.

Let $f_1^* \neq f_1$ be the funding spread required by the debt holders. The equivalence in equation (7) does not hold anymore (and so does not also the M&M theorem and its extension by Merton). The bank should use equation (6), where $f_1$ is set equal to the spread requested by the debt holders $f_1^*$, and the credit spread is the “enhanced” spread $cs_1^+$. 

---

Spring 2014

23
Formula (6) allows also to consider different risk-premia implicit in the preferences of the shareholders and bond holders: in perfect markets, in the simple framework we are working in, the risk-premia are reflected in the PD of the asset’s issuer. In the pricing process, if the bond holders use a different PD from the one used by shareholders (or by the bank’s managers, if they operate mainly in the interest of the latter), then the irrelevance of the leverage does not hold anymore and the correct formula to use. Different PDs are the consequence of different premia due to a limited information of the bond holders about the issuer of the asset bought by the bank.4

Including the Economic Capital

The economic capital is a concept used by financial institutions to indicate the amount of equity capital needed to keep the probability of default (i.e., of ending the business activity) below a given level, according to a model to measure the future uncertainty of the assets’ value.5

We can embed a premium in the return on equity (upon the risk-free rate) different from the one implicitly originated by the use of a PD containing a risk-premium fair to the bank. Let us indicate by \( \pi \) this additional return on equity and with \( EC_1 \leq E \) the amount of economic capital “absorbed” by asset \( A_1 \). Equation (5) modifies as follows:

\[
VB(t) = E \left[ X_1 \times (1 + ms_1) + E - X_1 \times (1 + f_1) \right] = (ms_1 - f_1)X_1 + E(1 - PD_1) + \\
+ \sum_{j=1}^{J} \max[X_jRec_j, + E - X \times (1 + f_1); 0]p_1PD_1 \tag{9}
\]

The margin spread including this additional premium is:

\[
ms_1 = \frac{E - \frac{X_1}{1 - PD_1}EC_1 \times X_1 \pi}{1 - PD_1} + f_1 = cs_1 + f_1 + cc_1 \tag{10}
\]

where \( cc_1 = \frac{EC_1}{1 - PD_1} \) is the cost of the economic capital associated to asset \( A_1 \).

**Proposition 2.** In internally evaluating an asset held by the bank, the economic capital enters in the bank’s internal pricing process through the return \( \pi \) requested by the shareholders.

Including Interest Rates Different from Zero

It is relatively easy to introduce non-zero interest rates in the framework sketched above: we limit the analysis to the case of a single, constant interest rate equal to \( r \). In this case equation (1) modifies as follows:

\[
A_1(t) = E \left[ \frac{1}{1 + r} X_1 \times (1 + r + s_1) \right] \\
= X_1 \times (1 + r + s_1)(1 - PD_1) + \frac{X_1Rec_1PD_1}{1 + r} = X_1 \tag{11}
\]

After same manipulations (see also Castagna and Fede [10], ch. 8), we have a formula for the credit spread \( s_1 \) similar to formula (2), where the loss given default is defined as \( Lgd_1 = 1 + r - Rec_1 \); for typical values of the interest rates, the two \( Lgd_1 \)'s do not differ too much.

Similarly, when rates are different from zero, the bank’s debt equation (3) reads:

\[
D_1(t) = E \left[ \frac{X_1 \times (1 + r + f_1)}{1 + r} \right] \\
= X_1 \times (1 + r + f_1)(1 - PD_1) + \\
+ \sum_{j=1}^{J} \min[X_jRec_j, + E(1 + r); X_1 \times (1 + r + f_1)]p_1PD_1 \\
= X_1 \tag{12}
\]

so that the funding spread has still a formula as in (4), with:

\[
Lgd_1 = 1 + r - \sum_{j=1}^{J} \min[X_jRec_j, + E(1 + r); X_1 \times (1 + f_1)]p_1PD_1
\]

Finally, the fair mark-up spread \( ms_1 \) the bank has to charge on asset \( A_1 \) when rates are different from zero, is given by solving the following equation:

\[
VB(t) = E \left[ \frac{X_1 \times (1 + r + ms_1) + E(1 + r) - X_1 \times (1 + r + f_1)}{1 + r} \right] \\
= \frac{[(ms_1 - f_1)X_1 + E(1 + r)](1 - PD_1)}{1 + r} \\
+ \sum_{j=1}^{J} \max[X_jRec_j, + E(1 + r) - X_1 \times (1 + f_1); 0]p_1PD_1 \\
= E \tag{13}
\]

\[\text{footnotes}
4It is quite reasonable to assume that typically banks have a deeper knowledge of their clients, which are ultimately the issuers of most of the assets. On the other hand, when the asset is a publicly traded security, it is more reasonable to assume that the information is evenly distributed amongst market participants. This is the case implicitly assumed in the M&M’s and Merton’s works.
5The level is often set by national and international regulation, such as the Basel regulation for banks.
\]
The formula for \(ms_1\) is similar as (6), if \(E(1+r) < X_1Lgd_{ij} + X_1f_1\) for one more \(Lgd_{ij}\)’s:

\[
ms_1 = \frac{E(1+r)-\overline{R}_jPD_1}{X_1(1-PD_1)} + f_1 = cs_1^* + f_1 \tag{14}
\]

where \(\overline{R}_j = \sum_{j=1}^{t} \max[X_1Rec_j + E(1+r) - X_1(1+f_1); 0]p_j\).

If \(E(1+r) < X_1Lgd_{ij} + X_1f_1\) in all cases, then \(ms_1 = (Lgd_{ij}PD_1)/(1-PD_1)\) with \(Lgd_{ij} = 1 + r - Rec_1\).

In conclusion, the introduction of interest rates into the analysis does not significantly affect results. Let the total yield of the asset \(A_1\) be \(i_1\): it can be written as the sum of the following components:

\[
i_1 = r + ms_1 = r + cs_1^* + f_1 + cc_1 \tag{15}
\]

We present now an example to illustrate in practical terms what we have stated in this section.

**EXAMPLE 1** Assume that the bank starts its activity in \(t = 0\) with an amount of equity capital \(E = 35\), which is deposited in a bank account \(B(0) = 35\). The bank wishes to invest in an asset whose price is \(A_1(0) = 100\) and it issues debt \(D_1(0) = 100\) to buy it. We will suppress the time reference in the following to lighten the notation. A sketched bank’s balance sheet is the following:

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B = 35)</td>
<td>(D_1 = 100)</td>
</tr>
<tr>
<td>(A_1 = 100)</td>
<td></td>
</tr>
</tbody>
</table>

\[E = 35\]

We assume interest rates are zero and that the issuer of the asset \(A_1\) can default with probability \(PD = 5\%\). Moreover, upon default, the recovery rate for the bank is stochastic: possible outcomes and the associated probabilities are in table 1. Clearly, \(Lgd_{ij} = 1 - Rec_1 = 60\%\). Firstly, let us determine which is the fair (credit) spread \(s_1 = cs_1\) requested by a non-leveraged investor. This is given by equation (2):

\[
s_1 = cs_1 = \frac{60\% \times 5\%}{1 - 5\%} = 3.158\%
\]

so that at the expiry the asset has a terminal value of \(A_1 = 100 \times (1 + 3.158\%)\). The bank is a leveraged investor, since it issues an amount of debt sufficient to buy the asset. Assuming we are in a market where perfect information is available to all participants, then the creditors of the bank know that it will buy the asset \(A_1\) and consequently they set a credit spread on the debt \(D_1\), which is a funding spread for the bank, by applying equation (4), by means of a two-step iterative procedure (the first step is computed by setting \(f_1 = 0\)), so that we get:

\[
f_1 = 1.406\%
\]

We can now compute the fair margin \(ms_1\) that the bank should charge on asset \(A_1\), by means of formula (6). We start with computing the different \(R_{1,5}\): they are shown in the table below. Since with these quantities we can compute also the bank’s default probability, by means of formula (8), this is shown in the table below as well.

<table>
<thead>
<tr>
<th>(R_{1,5})</th>
<th>(PD_B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.719</td>
<td>0.000%</td>
</tr>
<tr>
<td>-</td>
<td>3.500%</td>
</tr>
<tr>
<td>-</td>
<td>0.500%</td>
</tr>
</tbody>
</table>

\[1.719 \times 4.000\%\]

We have all we need to calculate the “adjusted” credit spread \(cs_1^*\):

\[
 cs_1^* = \frac{35 - 1.295\%}{100 - 5\%} = 1.752\%
\]

which plugged in (6):

\[
 ms_1 = cs_1^* + f_1 = 1.752\% + 1.406\% = 3.158\% = s_1
\]

thus confirming (7).

The equity should be used to prevent a default of the bank occurring with a probability higher than a given level. The economic capital is the amount of equity necessary to keep the probability below the targeted level. In this example, assuming that the target probability is 4\%, we can see from the calculations above that actually \(PD_B = 4\%\).
It is easy to prove that the amount of economic capital to achieve this target is just above 26.75, not 35, so EC1 = 26.75 and the remaining is equity capital that could be invested in other risky assets. We suppose that the stockholder requires a return on economic capital of \( \pi = 5\% \). The cost of capital is:

\[
c_{1} = \frac{26.75 \times 5\%}{1 - 5\%} = 1.408\%
\]

The total \( m_{1} \) to charge on \( A_{1} \), by formula (10), will then become

\[
m_{1} = cs_{1} + f_{1} + c_{1} = 1.752\% + 1.406\% + 1.408\% = 4.556\%
\]

### Balance Sheet with Two Assets (Uncorrelated Defaults)

We now move on to a multi-period setting, where the bank, after the investment in asset \( A_{1} \), decides to further invest in a new asset \( A_{2} \), whose initial price is \( X_{2} \) and terminal pay-off \( A_{2}(T) = X_{2} \times (1 + s_{2}) \). Without a great loss of generality, we assume that the expiry of the asset \( A_{2} \) is the same as asset \( A_{1} \), in \( T \), and that the investment in the new asset occurs in \( t^{+} \), just an instant after the initial time \( t \); to lighten the notation, we will set \( t^{+} = t \) in what follows, even though they are two distinct instants.

Also for asset \( A_{2} \) there is a probability \( PD_{2} \) that the asset’s issuer defaults, in which case the buyer of the asset will receive a stochastic recovery \( R_{2} \) expressed as a percentage of the face value \( X_{2} \). \( R_{2} \) can take values \( R_{2,l} \), for \( l = 1, \ldots, L \), and each possible value can occur with probability \( p_{2,l} \). We assume for the moment that the defaults of the issuers of assets \( A_{1} \) and \( A_{2} \) are uncorrelated.

Similarly to asset \( A_{1} \), for a (non-leveraged) investor

\[
A_{2}(t) = E[X_{2} \times (1 + s_{2})] = X_{2} \times (1 + s_{2})(1 - PD_{2}) + X_{2}R_{2}PD_{2} = X_{2}
\]

where \( R_{2} = \sum_{l=1}^{L} R_{2,l}p_{2,l} \) is the expected recovery in the event of the issuer’s bankruptcy. The fair credit spread is derived as n the case of asset \( A_{1} \) and it is:

\[
s_{2} = cs_{2} = \frac{Lgd_{2}PD_{2}}{1 - PD_{2}}
\]

where \( Lgd_{2} = (1 - R_{2}) \).

Assume now that the bank buys the asset \( A_{2}(t) = X_{2} \) by issuing new debt: the total debt is \( D(t) = D_{1}(t) + D_{2}(t) = X_{1} + X_{2} = X \), i.e.: the leverage increases as well. Let \( f_{2} \) be the funding spread paid on debt \( D_{2}(t) \).

The funding spread requested by bank’s bond holders on the new debt \( D_{2}(t) \) is derived in a way similar to equation (3):

\[
D_{2}(t) = E[X_{2}(1 + f_{2})] = X_{2}(1 + f_{2})(1 - PD_{1})(1 - PD_{2}) + \sum_{l=1}^{L} \min \left[ \frac{X_{2}R_{1} + X_{2}(1+s_{2}) + E}{X}; (1 + f_{2}) \right] \\
* X_{2}p_{1,PD_{1}}(1 - PD_{2}) + \sum_{l=1}^{L} \min \left[ \frac{X_{2}(1+s_{2}) + X_{2}R_{2} + E}{X}; (1 + f_{2}) \right] (18)
\]

\[
* X_{2}p_{2,PD_{2}}(1 - PD_{1}) + \sum_{l=1}^{L} \sum_{l=1}^{L} \min \left[ \frac{X_{2}R_{1} + X_{2}R_{2} + E}{X}; (1 + f_{2}) \right] * X_{2}p_{1,PD_{1}}(1 - PD_{2})PD_{2}PD_{1} = X_{2}
\]

The spread \( s_{2} \) can be either the credit spread or the mark-up spread set by the bank. The funding spread \( f_{2} \) can be found by solving equation (18): also in this case, as for the case in the previous section for a single asset balance sheet, the solution \( f_{2} \) is found by a very quick numerical search, starting by setting \( f_{2} = 0 \). It should be noted that in this stylised multi-period setting, the funding spread \( f_{1} \) on the first debt \( D_{1} \) enters in the pricing of the new debt \( D_{2} \) as an input: the bank cannot change this parameter that is contractually set until the expiry in \( T \) and this is consistent with the real world. It should be also noted that this spread is no more fair and it should be revised, since the new recovery ratio for the total debt \( D \) is now different from the original one set before the investment in the asset \( A_{2} \). This anyway cannot happen, as the outstanding debt pays a contract rate fixed until the expiry.

The mark-up spread, set by the bank on the second asset, is such that the expected net value of the bank is still the amount of equity posted by shareholders:

\[
VB(t) = E[X_{1}(ms_{1} - f_{1}) + X_{2}(ms_{2} - f_{2}) + E] = [X_{1}(ms_{1} - f_{1}) + X_{2}(ms_{2} - f_{2}) + E] * (1 - PD_{1})(1 - PD_{2}) + \sum_{j=1}^{L} \max \left[ X_{1}(R_{1} - (1 + f_{1})) + X_{2}(ms_{2} - f_{2}) + E; 0 \right] * p_{1,PD_{1}}(1 - PD_{2}) + \sum_{j=1}^{L} \max \left[ X_{1}(R_{1} - (1 + f_{1})) + X_{2}(ms_{2} - f_{2}) + E; 0 \right] * p_{2,PD_{2}}(1 - PD_{1}) + \sum_{j=1}^{L} \sum_{k=1}^{L} \max \left[ X_{1}(R_{1} - (1 + f_{1})) + X_{2}(ms_{2} - f_{2}) + E; 0 \right] * p_{1,PD_{1}}(1 - PD_{2})PD_{2}PD_{1} = E
\]
Let
\[
\mathcal{R}_1^* = \sum_{j=1}^{l} \max \left[ X_1(\text{Rec}_1) - (1 + f_1) \right] + X_2(ms_2 - f_2) + E; 0p_1, \]
\[
\mathcal{R}_2^* = \sum_{i=1}^{l} \max \left[ X_i(ms_1 - f_1) \right] + X_2(\text{Rec}_2) - (1 + f_2) + E; 0p_2, \]
and
\[
\mathcal{R}_{1,2}^* = \sum_{j=1}^{l} \sum_{i=1}^{l} \max \left[ X_i(\text{Rec}_i) - (1 + f_1) \right] + X_2(\text{Rec}_2) - (1 + f_2) + E; 0p_1, p_2. \]

We will rewrite the (19) in a lighter notation:
\[
\mathcal{V}_B(t) = E \left[ X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E \right] = \left[ X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E \right] \times (1 - PD_1)(1 - PD_2)
\]
\[
+ \mathcal{R}_1^* PD_1(1 - PD_2) + \mathcal{R}_2^* PD_2(1 - PD_1)
\]
\[
+ \mathcal{R}_{1,2}^* PD_1 PD_2 = E \tag{20} \]

The solution $ms_2$ to (20) can be found by a quick numerical procedure, by initialising $ms_2 = 0$. It is possible, anyway, to distinguish two cases that will shed some light on which is the value that $ms_2$ can take.

Before analysing the two distinct cases, we provide the probability of default of the bank once the new asset is bought. As before, we need to check for all the cases when the bank’s value drops below zero, in which case the limited shareholders’ liability floors the value at zero. We have that:
\[
PD_B = \sum_{j=1}^{l} \sum_{i=1}^{l} 1(\mathcal{R}^*_i = 0) p_1 PD_1(1 - PD_2)
\]
\[
+ \sum_{i=1}^{l} 1(\mathcal{R}^*_i = 0) p_2 PD_2(1 - PD_1) \tag{21}
\]
\[
+ \sum_{j=1}^{l} \sum_{i=1}^{l} 1(\mathcal{R}^*_i = 0) p_1 p_2 PD_1 PD_2
\]

where $1_{\{\cdot\}}$ is the indicator function and $\mathcal{R}^*_i$ is the $b$-th addend in the summation in each $\mathcal{R}^*_i$.

The Default of the Asset $A_2$ Does not Imply the Default of the Bank

First, we assume that the default of the asset $A_2$ does not imply the default of the bank. This may happen for a variety of reasons, but mainly because the quantity $X_2$ of the asset bought by the bank is small compared to the entire balance sheet. So, even in the event of bankruptcy of the issuer of $A_2$, the bank is able to cover losses and to repay all its creditors without completely depleting its equity capital $E$ and hence declaring default. On the other hand, the default of the asset $A_1$ still causes the default of the bank as before.

If $X_2$ is much smaller than the quantity $X_1$ of asset $A_1$ already included in the assets of the bank’s balance sheets, then we have the following approximations:
\[
\mathcal{R}_1^* \approx \mathcal{R}_{1,2}^* \approx \mathcal{R}_1
\]
and
\[
\mathcal{R}_2^* \approx X_1(ms_1 - f_1) + X_2(\text{Rec}_2 - (1 + f_2)) + E
\]

After substituting the values above, equation (20) can be written as:
\[
\mathcal{V}_B(t) = E \left[ X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E \right]
\]
\[
\times (1 - PD_1)(1 - PD_2)
\]
\[
+ \mathcal{R}_1^* PD_1(1 - PD_2) + \mathcal{R}_2^* PD_2(1 - PD_1)
\]
\[
+ \mathcal{R}_{1,2}^* PD_1 PD_2 = E \tag{22}
\]

Let $[(ms_1 - f_1) X_1 + E] = H; we can re-write (22) as
\[
\mathcal{V}_B(t) = H[(1 - PD_1)(1 - PD_2) + PD_2(1 - PD_1)]
\]
\[
+ \mathcal{R}_1^* PD_1 + [X_2(ms_2 - f_2)] (1 - PD_1)(1 - PD_2)
\]
\[
+ [X_2(\text{Rec}_2 - (1 + f_2))] PD_2(1 - PD_1) \tag{23}
\]
\[
= E
\]

By formula (5) and the definition of $ms_1$:
\[
[(ms_1 - f_1) X_1 + E](1 - PD_1) + \mathcal{R}_1^* PD_1 = E
\]

then the first term on the RHS of the first line in (23) is:
\[
H[(1 - PD_1)(1 - PD_2) + PD_2(1 - PD_1)] + \mathcal{R}_1^* PD_1
\]
\[
= H(1 - PD_1) + \mathcal{R}_1^* PD_1 = E
\]

so that:
\[
\mathcal{V}_B(t) = [X_2(ms_2 - f_2)] (1 - PD_1)(1 - PD_2) + \mathcal{R}_1^* PD_1 = 0 \tag{24}
\]

Recalling that $X_2(1 - \text{Rec}_2) = X_2 Lgd_2$, equation (24) simplifies:
\[
\mathcal{V}_B(t) = [X_2(ms_2 - f_2)] (1 - PD_2) - (X_2 Lgd_2 + f_2) PD_2(1 - PD_1) = 0 \tag{25}
\]

which holds if
\[
[X_2 ms_2 (1 - PD_2) - X_2 f_2 - X_2 Lgd_2 PD_2] = 0
\]
or
\[
ms_2 = \frac{Lgd_2 PD_2 + f_2}{1 - PD_2} = cs_2 + \frac{f_2}{1 - PD_2} \tag{26}
\]

The fair margin spread $ms_2$ is no more equal to the credit spread $cs_2$, as it was the case for asset $A_1$ (see equation (7)). It is worth noting that the total margin spread is made of two components:
the credit spread $cs_2$, equal to the spread requested by a non-leveraged investor. This is different from the credit spread required on asset $A_1$ by the bank, in equation (6). The reason is that when the asset $A_1$ issuer goes defaulted, this will cause also the the bank’s default and the shareholders are liable only up to the amount $E$ to the debt holders, hence a reduced credit spread $cs'_1$ can be applied; but the default of the issuer of the asset $A_2$ does not cause the bank’s default, so that the loss is fully covered by the equity, so in this case the limitation of the liability up to $E$ is immaterial, and the entire (expected) loss given default $Lgd_2$ must be born by the shareholders. As a consequence, the credit spread applied on the asset $A_2$ in the mark-up margin is equal to $cs_1$, in the non-leveraged investor case;

- the funding spread $f_2$, conditioned to the survival of the issuer of the asset $A_2$. Also in this case there is a difference with the case of the asset $A_1$: the funding spread in (6) is not divided by the survival probability of the issuer of $A_1$ since the bank is interested in receiving the amount $f_2 X_1$ without considering the issuer’s default. In fact, when the issuer of $A_1$ goes defaulted, the margin spread $m_1$ is not cashed in by the bank and it cannot repay the funding costs on the debt $D_1$. But in the event of default also the bank goes bankrupt, so that shareholders are not interested in the missing payments since they are anyway limited up to the amount $E$. In case of asset $A_2$, the missing payment of $f_2$ has to be covered by the shareholders, which will pay $f_2 X_2$ on the debt $D_2$ in any case. The expected loss due to this cost must equal the expected return embedded within the mark-up spread $ms_2$:

$$E[f_2 X_2] = f_2 X_2 (1 - PD_2) = f_2 X_2,$$

which means that $f_2 = f_2 / (1 - PD_2)$.

We can state the following result:

**Proposition 3.** When pricing an asset that represents a small percentage of the bank’s total assets and whose default does not affect the bank’s default, the correct (approximated) and theoretically consistent mark-up margin to apply includes the issuer’s credit spread, fair to a non-leveraged investor, plus the bank’s funding spread conditioned to the issuer’s survival probability.

This result is an accordance with the result in Castagna [5], where the fair spread that an issuer would pay to a non-leveraged investor includes only the credit spread ($cs_2$ in this case), whereas the mark-up margin charged by the leveraged bank includes also the funding costs (see also chapter 11 in Castagna and Fede [10]). In Castagna [5] the irrelevance of the default of the bank was postulated in an axiom based on the going concern principle. In this work on the contrary we are explicitly considering the case when such irrelevance arises. An example is given by an asset representing a small percentage of the total bank’s assets: when its issuer defaults, this bankruptcy would not affect the survival of the bank. In this case the results are re-derived and confirmed in a much theoretically sounder framework extending the simplified economies in Modigliani& Miller and Merton, thus delimiting the cases when their results remain fully applicable.

The results just shown confirm also that the practice of including the funding valuation adjustment (FVA) in the valuation (i.e.: internal pricing) process of a contract is fully justified: this thesis was supported in Castagna [6] (arguing against the opposite view in Hull&White [13] and [14]) but not proved analytically. A first attempt to extend the Modigliani&Miller [18] and Merton [16] framework in a multi-period setting was in Castagna [9], but there the simplified structure of the bank’s balance sheet did not allow to achieve a fully convincing justification of the all-inclusive pricing rules followed by practitioners. We here showed that the main point is to consider also a richer balance sheet with a more complex set of interrelations between assets’ issuers and the bank.

If we consider also the economic capital that the asset $A_2$ entails to ensure that the default probability of the bank is equal or smaller than given level, the margin spread can be easily modified:

$$ms_1 = \frac{Lgd_2 PD_2 + EC_2 / X_2 \pi + f_2}{1 - PD_2} = cs'_2 + \frac{f_2}{1 - PD_2} + cc_2$$

(27)

where $EC_2$ is the economic capital required by the investment in asset $A_2$ and $cc_2 = EC_2 / X_2 \pi$ is the associated cost.

**EXAMPLE 2** We reprise the Example 1 and we assume that an instant after the start of the activities in $t = 0$, i.e.: in $t'$ (which, by an approximation, we set equal to $t$) the bank invests in an asset $A_2$, whose issuer can default between times 0 and 1 with probability $PD_2 = 6\%$. There are three recovery scenarios, with associated probability, as shown in the following table:
We can immediately compute the fair spread $s_2 = cs_2$ that a non-leveraged investor would require on this asset:

$$s_2 = cs_2 = \frac{60\% \times 5\%}{1 - 6\%} = 3.83\%$$

The price in $t = t = 0$ is $A_2(0) = 10$ and the bank issues new debt for an equivalent amount $D_2(0) = 10$: the amount represents a relatively small fraction of the entire assets held by the bank; moreover, when the issuer of $A_2$ defaults, the bank will not go bankrupt in any recovery scenario. As in the first example, we suppress all references to time from now on. The bank’s balance sheet will now be:

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B = 35$</td>
<td>$D_1 = 100$</td>
</tr>
<tr>
<td>$A_1 = 100$</td>
<td>$D_2 = 10$</td>
</tr>
<tr>
<td>$A_2 = 10$</td>
<td>$E = 35$</td>
</tr>
</tbody>
</table>

The spread requested by the creditors of the bank can be found be recursively solving equation (18), starting by setting $f_2 = 0$. We come up with the result:

$$f_2 = 1.318\%$$

The fair spread $ms_2$ that the bank has to charge on asset $A_2$, can be found by means of (26)

$$ms_2 = cs_2 + \frac{f_2}{1 - PD_2} = 3.830\% + \frac{1.318\%}{1 - 6\%} = 5.148\%$$

The simple rule often followed by banks to set the spread (under the hypothesis that interest rates are zero) is to charge in the margin the credit spread referring to the issuer of the asset and the funding cost:

$$ms_2 = 3.830\% + 1.318\% = 5.148\%$$

in case of low issuer’s probability of defaults this is an approximation of the fair margin that should be applied. The probability of default of the bank, after the investment in the asset $A_2$, is calculated from equation (50) and it is $PD_2 = 4.01\%$. This means that the equity is almost fully sufficient to keep the probability of default at the chosen level of 4% we mentioned in example. To avoid unneeded complication, we take the new $PD_2$ as compliant with the limit in practice, which means that the economic capital absorbed by the new asset is $EC = 8.25$. If we want to include also the remuneration for $EC_2$ in the margin, then from (27) we have:

$$cc_2 = \frac{8.25 \times 5\%}{1 - 6\%} = 4.145\%$$

so that the total margin is $ms_2 = 5.148\% + 4.125\% = 9.273\%$. In this example the economic capital requested to keep $PD_2 \leq 4\%$ is equal to amount invested in $A_2$: this is due to the simplified balance sheet we are considered and to the small number of scenarios for the recovery of the two assets. In the real world, in a much more complex balance sheet, the economic capital for an additional investment is typically smaller than the invested amount and consequently also the cost of capital component of the margin is much smaller.

**The Default of the Asset $A_2$ Implies the Default of the Bank**

Assume now that the default of the asset $A_2$ implies the default of the bank: this is typically the case when the asset $A_2$ is, in percentage terms, a great share of the bank’s total assets. The margin spread that the bank has to apply on the second asset is derived numerically by solving equation (23). It is possible to write down a formally closed form formula, to assess which are the parameters affecting the spread.

By rewriting equation (20), we have:

$$VB(t) = E [X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E]
= [X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E]
\times (1 - PD_1)(1 - PD_2)
\times E$$

where $e = \bar{R}_1PD_1(1 - PD_2) + \bar{R}_2PD_2(1 - PD_1) + \bar{R}_2PD_2PD_1$. Let $PD_{1,2} = 1 - (1 - PD_1)(1 - PD_2)$; we solve for $ms_2$ and we get:

$$ms_2 = \frac{E \times \bar{X}_1PD_{1,2}}{(1 - PD_1)(1 - PD_2)} + f_2
\times \frac{X_1(ms_1 - f_1)(1 - PD_1) - \bar{R}_1PD_2}{(1 - PD_1)(1 - PD_2)} - f_2 - db$$

The margin spread $ms_2$ resembles the margin spread $ms_1$ derived for asset $A_1$: it is made of three constituent parts:

- an “adjusted” credit spread $cs_2^*$, due to the limited liability of the shareholders, referring to the asset $A_2$: the loss given default suffered by the bank, $E/X_2$, is smaller than the loss given default of the asset $A_2$, $Lgd_x$, suffered by a non-leveraged investor. This lower loss
<table>
<thead>
<tr>
<th>$X_2$</th>
<th>$f_2$</th>
<th>$m_{S_2}$</th>
<th>$db$</th>
<th>$PD_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.318%</td>
<td>5.229%</td>
<td>-0.081%</td>
<td>4.01%</td>
</tr>
<tr>
<td>20</td>
<td>1.215%</td>
<td>5.108%</td>
<td>-0.063%</td>
<td>4.05%</td>
</tr>
<tr>
<td>30</td>
<td>1.128%</td>
<td>5.007%</td>
<td>-0.050%</td>
<td>4.05%</td>
</tr>
<tr>
<td>40</td>
<td>1.059%</td>
<td>4.904%</td>
<td>-0.015%</td>
<td>4.63%</td>
</tr>
<tr>
<td>50</td>
<td>1.038%</td>
<td>4.762%</td>
<td>0.105%</td>
<td>4.63%</td>
</tr>
<tr>
<td>60</td>
<td>1.135%</td>
<td>4.570%</td>
<td>0.395%</td>
<td>8.62%</td>
</tr>
<tr>
<td>70</td>
<td>1.281%</td>
<td>4.274%</td>
<td>0.837%</td>
<td>8.62%</td>
</tr>
<tr>
<td>80</td>
<td>1.411%</td>
<td>4.073%</td>
<td>1.168%</td>
<td>8.62%</td>
</tr>
<tr>
<td>90</td>
<td>1.527%</td>
<td>3.932%</td>
<td>1.425%</td>
<td>8.62%</td>
</tr>
<tr>
<td>100</td>
<td>1.632%</td>
<td>3.832%</td>
<td>1.630%</td>
<td>8.62%</td>
</tr>
</tbody>
</table>

**TABLE 2**: Funding spread, margin, diversification benefit and bank’s probability of default as the amount invested in asset $A_2$ increases.

The given default multiplies the probability that either the issuer of the asset $A_2$ or the issuer of asset $A_1$ goes bust, to get the expected loss, and then divided by the survival probability of the bank (i.e.: the joint probability that both issuers of assets $A_1$ and $A_2$ survive). This means that the margin spread should equal the expected loss given default, given the bank’s survival;

- the funding spread $f_2$ paid by the bank on the new debt issued to buy the asset $A_2$. The funding spread is not divided by the survival probability of issuer since when it defaults, the bank will go bankrupt as well and it will not pay the spread to its creditors.

- a diversification benefit $db$: all the quantities entering in $db$ are positive, so that its effect on the margin is to abate $m_{S_2}$. Basically $db$ accounts for the fact that a part of the terminal value of the equity is made by the terminal value of $A_1$, provided that its issuer does not default; moreover, when the default of either issuers, or of both, occurs then the total recovery value for the bank considers both terminal values $A_1$ and $A_2$ (the quantity $c$). These two quantities are divided by the survival probability of the bank (i.e.: the joint probability that both issuers of assets $A_1$ and $A_2$ survive).

Even if the capital markets are perfect, the total margin spread $m_{S_2}$ differs from the market credit spread $c_{S_2}$ and it will typically be greater than the latter. When the amount of asset $A_2$ held by the bank increases, the diversification benefit $db$ tends to abate the margin $m_{S_2}$ and it can be also lower than $c_{S_2}$ for sufficiently large $A_2$. In any case $m_{S_2}$ will not be the sum of the credit spread $c_{S_2}$ and the bank’s funding spread $f_2$, as in the case above when the default of the asset $A_2$ did not imply the bank’s default.

**Proposition 4.** When evaluating an asset whose default implies the bank’s default, the correct and theoretically consistent mark-up margin to apply includes and “adjusted” issuer’s credit spread, the funding spread paid of the new debt to buy the asset, and a diversification benefit due to the different recoveries of assets and on the non-simultaneous defaults, since they are uncorrelated.

**EXAMPLE 3** We continue from the Example 1 and we assume that the bank wishes to invest in $A_2$ an amount that is big enough to trigger its own default when the issuer goes bust, in at least one recovery scenario. In this case formula (29) is computed by numerically solving (20) by a numerical procedure. In table 2 we show the effect of increasing the investment in $A_2$ by issuing new debt. The probability of default will increase because the shareholder would pour more equity in the bank to preserve the desired target level (it was set at 4% before). Consequently, also the funding spread $f_2$ increases, since the expected losses suffered by bank’s creditors will be higher due to the higher $PD_B$ and the lower recoveries. The diversification benefit will be negative for relatively small amount of $A_2$. We have also a confirmation that for small amounts of $A_1$, as $X_2 = 10$ of the example before) formula (26) is a good approximation, since we are getting the same result 5.229%. For larger amounts of $A_2$ the benefit becomes positive, thus lowering the margin applied on the asset. The margin will approach the level required by a non-leveraged investor for $A_2 = 100$. The diversification benefit $db$ starts being positive when the amount of $X_2 > 40$, or

---

6When the default of $A_1$ or $A_2$ can both trigger the default of the bank, the loss is suffered when either of the two issuers go bankrupt, whence the weighting by the probability $PD_{1,2}$. 
when the default of the issuer of $A_2$ implies, in at least one recovery scenario, the default of the bank.

Balance Sheet with Two Assets (Perfectly Correlated Defaults)

Assume now that the defaults of the issuers for assets $A_1$ and $A_2$ are perfectly correlated: let $1_{A_1}$ the indicator function equal to 1 when the default of the issuer of $A_1$ occurs, and $1_{A_2}$ similarly defined. Moreover let $\Pr[D]$ be the probability of default of both issuers. We have:

$$\Pr[1_{A_1} = 1 \cap 1_{A_2} = 1] = \Pr[D]$$
$$\Pr[1_{A_1} = 0 \cap 1_{A_2} = 0] = 1 - \Pr[D]$$
$$\Pr[1_{A_1} = 1 \cap 1_{A_2} = 0] = \Pr[1_{A_1} = 0 \cap 1_{A_2} = 1] = 0$$

The fair funding spread requested by the creditor of the bank on the new debt $D_2$, in a perfect market, is derived by solving the following equation:

$$D_2(t) = E[X_2(1 + f_2)] = X_2(1 + f_2)(1 - \Pr[D])$$
$$+ \sum_{j=1}^{l} \sum_{i=1}^{l} \min \left[ \frac{X_j \text{Rec}_j + X_i \text{Rec}_i - E}{X} ; (1 + f_2) \right] X_j p_1 p_2 \Pr[D]$$
$$= X_2$$

(30)

The margin the bank has to apply on the asset $A_2$, provided it has enough bargaining power, is obtained by a modified version of equation (31) that accounts for the perfect correlation between the issuers of the two assets:

$$\text{VB}(t) = E[X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E]$$
$$= [X_1(ms_1 - f_1) + X_2(ms_2 - f_2) + E](1 - \Pr[D])$$
$$+ \overline{X}_{1,2} \Pr[D] = E$$

(31)

where the notation is the same used before.

If we further assume that the bankruptcy of the two issuers triggers, in at least in one of the possible cases of loss given default, the default of the bank then the formula for the fair spread $ms_2$ can be formally written as:

$$ms_2 = \frac{f_2}{(1 - \Pr[D])} + f_2 - \frac{X_2(\text{ms}_1 - f_1)(1 - \Pr[D]) + \overline{X}_{1,2} \Pr[D]}{(1 - \Pr[D])}$$

(32)

Formula (32) is similar to (29) for the case when default are correlated. The main difference is that the joint probability of default of the two issuers $(1 - \Pr[D_1])(1 - \Pr[D_2])$ has been replaced by $(1 - \Pr[D])$, the probability of the issuer of asset $A_2$ is $\Pr[D]$, and in the diversification benefit $db$, the quantity $\epsilon$ is replaced by a smaller amount $\overline{R}_{1,2}$: this is due to the impossibility to exploit the value of the non defaulting asset to cover the losses of the defaulting one, since this situation can never occur if the two defaults are correlated.

**Proposition 5.** When pricing an asset whose default implies the bank’s default and that is perfectly correlated with the default of the other asset(s) held by the bank, the correct and theoretically consistent mark-up margin to apply includes “adjusted” issuer’s credit spread, the funding spread paid of the new debt to buy the asset, and a limited diversification benefit due to the different possible recoveries on assets whose defaults are correlated.

Balance Sheet with a Derivative Contract

We have analysed how to bank should evaluate (i.e.: internally price) assets when it decides to invest in them. Assume now that the bank, after investing at time $t = 0$ in asset $A_1$, enters in a derivative contract, instead of buying another asset. Actually, entering in a contract can be seen as buying an asset or issuing a liability, or in some cases doing both depending on contingent evolution of (typically financial) variables (as for example when a swap contract is closed). In this case, the valuation of the contract can be operated as it were an asset (or a liability) with stochastic intermediate cash flows and terminal pay-off, whose expected amounts are discounted at the reference date.

We would like here to stress that a bank can also offer a service of market-making for derivative contracts: in this case it does not really try to buy an asset or issue a liability; on the contrary, the bank is selling a product manufactured with a given technology. For example, the bank can be a market-maker for options on a given asset: it is not really interested in buying or selling options based on some expectation of the future evolution of the underlying asset, simply it sells a product that can be a long or a short position in the option. Once one of two positions is sold to the client, the bank has the internal skills to use the available technology to manufacture the product so that its profit derives from the ability to sell at a margin over the production cost.7

7In Castagna [9] we adopt this “industrial” metaphor in a rather discursive argument to justify the inclusion of funding cost in the internal evaluation of contracts by the bank. In this work we will expand in a more quantitative way the ideas presented in there.
satisfies the following conditions:

1. Self financing condition, that is: no other investment is required in operating the strategy besides the initial one;
2. Replicating condition, that is: at any time \( t \) the replicating portfolio’s value equals the value of the contract.

Keeping the analysis in a discrete time setting, we work in the classical binomial setting by Cox and Rubinstein [12]. Assume that the underlying asset is \( S(t) = S \) at time \( t \), and that it can go up to \( S_u = Su \) or down to \( S_d = S \), with \( d < 1, u > 1 \) and \( u \times d = 1 \) in next period \( T \), with a probability, respectively, equal to \( q \) and \( 1 - q \). Let \( V(t) \) be the price of a derivative contract at time \( t \), and \( V_u \) and \( V_d \) its value when the underlying jumps to, respectively, to \( S_u \) and \( S_d \). Besides, let \( B(t) = B \) be the value in \( t \) of a risk-free bank-account deposit earning the risk-free rate. We want to replicate a long position in the derivative contract.

To do so, we have to set the following equalities in each of the two state of the world (i.e.: possible outcomes of the underlying asset’s price):

\[
V_u = aSu + \beta B(1 + r) \tag{33}
\]

and

\[
V_d = aSd + \beta B(1 + r) \tag{34}
\]

Equations (33) and (34) are a system that can be easily solved for quantities \( a \) and \( \beta \), yielding:

\[
a = \Delta = \frac{V_u - V_d}{u - d} \tag{35}
\]

and

\[
\beta = \frac{uV_d - dV_u}{(u - d)B(1 + r)} \tag{36}
\]

We have indicated \( a = \Delta \) because it is easily seen in (35) that it is the numerical first derivative of the price of the contingent claim with respect to the underlying asset, usually indicated so in the Option Pricing Theory.

If the replicating portfolio is able to mimic the pay-off of the contract, then its value at time \( t \) is also the arbitrage-free price of the contract:

\[
V(t) = \Delta S + \beta B = \frac{V_u - V_d}{(u - d)} + \frac{uV_d - dV_u}{(u - d)(1 + r)} \tag{37}
\]

It is possible to express (37) in terms of discounted expected value under the risk neutral measure:

\[
V(t) = \frac{1}{1 + r} [pV_u + (1 - p)V_d] \tag{38}
\]

with \( p = \frac{(1 + r) - d}{u - d} \). \( V(t) \) is the expected risk neutral terminal value of the contract, discounted with the risk-free rate.

This is the original setting by Cox and Rubinstein [12] that assumed a perfect and friction-less market, where default risk is absent (hence, the bank cannot go bankrupt and there are no differential rates for borrowing or lending). But if the bank is a defaultable agent, it is no more guaranteed that the replication argument works exactly as described above. We try to investigate this matter now.

Let us go back to our setting above, after the bank started its activities and invested in asset \( A_1 \), at time \( t’ \). As before, for sake of simplicity of the notation, we set \( t’ = t \) in practice, even if they are two distinct instants. At this time, the bank closes a derivative contract denoted by \( \phi V \), where \( \phi = \pm1 \) depending on whether the bank has, respectively, a long or short position in it. The replication strategy involves trading in a quantity \( -\phi \alpha \) in the underlying asset \( S \) and \( -\phi \beta \) in a bank deposit: in the real world being long a risk-free bank account means for the bank lending money to a risk-free counterparty\(^8\) whereas being short means for the bank borrowing money.

The standard replication argument does not consider the default of the replicator (the bank) so that borrowing and lending money occurs at the same interest rate. In reality, the replicator can borrow money at a rate which possibly includes a spread for the risk of its own default: this is the funding spread that we analysed above.\(^9\) In the following analysis we assume that the default of the counterparty of the derivative contract (occurring with probability \( PD_V \)) is independent from the evolution of the underlying asset and that the issuer of the latter cannot default. The possible recoveries of the final pay-off of the derivative contract are \( \text{Rec}_l \), for \( l = 1, \ldots, L \), each with occurring with probability \( p_{lV} \).

First we need to know if the bank is able to borrow money (if this is prescribed by the replication strategy) at the risk-free rate, as supposed by the theory. If the bank needs to borrow money (i.e.: \( -\phi \beta < 0 \), it will issue new debt \( D_2 = |\phi \beta B| \); in a market with perfect information the creditors will set a credit (funding, from bank’s perspective) spread to remunerate the credit risk. If the bank has to invest positive cash-flows in a risk-free bank

\(^8\)It can be alternatively assumed that money is lent to a defaultable counterparty, at a rate including a credit spread compensating for the risk. In this case the expected interest rate yielded by the loan is the risk-free rate, when accounting for the expected losses given default.

\(^9\)See also Castagna [8] and Castagna and Fede [10] for more details on the replication of contract including funding and liquidity costs. For an extended formal treatment of the replication of a contract in a world with differential rates, see Mercurio [17].
account (i.e. $-\phi \beta > 0$), it will earn the risk-free rate (set equal to 0 in the current analysis).

The credit spread required by the creditor of the bank (alternatively said, the funding spread $f_2$ the bank has to pay) is determined by the following equation (recall we are setting $r = 0$), the extension to non-zero rates is straightforward:

\[
D_2(t) = E[|\phi \beta B| (1 + f_2)] = [\phi \beta B](1 + f_2)(1 - PD_1)(1 - PD_V)
\]

\[
+ \sum_{j=1}^{J} \min \left[ \frac{X_{i} \text{Rec} + \phi \left( V_T - \Delta S \right) + E}{X_{i} + |\phi \beta B|}; (1 + f_2) \right] |\phi \beta B|
\]

\[
* p_1 PD_1 (1 - PD_V)
\]

\[
+ \sum_{j=1}^{J} \min \left[ \frac{X_{i} (1 + s_1) + \text{EPE}_j \text{Rec} - \phi \Delta S + E}{X_{i} + |\phi \beta B|}; (1 + f_2) \right] (39)
\]

\[
* |\phi \beta B| p_2 PD_V (1 - PD_1)
\]

\[
+ \sum_{j=1}^{J} \sum_{l=1}^{L} \min \left[ \frac{X_{i} \text{Rec} + \text{EPE}_j \text{Rec} - \phi \Delta S + E}{X_{i} + |\phi \beta B|}; (1 + f_2) \right]
\]

\[
* |\phi \beta B| p_1. p_V. PD_1. PD_1
\]

where \( \text{EPE}_V = \phi V^+ (T) = \phi [p V_{i} I_{\{\phi V_{i} > 0\}} + (1 - p) V_{i} I_{\{\phi V_{i} < 0\}}] \) is the expected positive exposure (the bank looses on the counterparty’s default only if the contract has a positive value); \( V = E[V] \) and \( \Sigma = E[S] \). Basically the first second line of equation (39) is the amount the creditors of the bank receive if the bank survives (i.e. if no default of the issuer of asset \( A_1 \) or counterparty of the derivative contract) occurs; the third, fourth and fifth lines is the amount that the creditors of the bank expect to recover if, respectively, the issuer of asset \( A_1 \), the counterparty of the derivative contract or both default. The total expected amount returned to the creditors equals the present amount borrowed by the bank, in the last line.

As we have seen above when adding a new asset \( A_2 \), if we focus on the case when the new contract does not affect the default of the bank, than we should be able to exclude also the cases when the residual value of the assets is negative. Although this cannot happen when replicating plain vanilla options, it cannot be excluded for some exotic options. We will not deal with this very specific situations, but they can be dealt with by introducing a floor at zero in the arguments of the \( \min \) functions.

It should be also manifest that the funding spread \( f_2 \) is above zero for non trivial cases, even assuming a perfect market. This is due to the fact that the composition of the existing balance sheet, when closing the contract and starting its replication, does affect the ability of the bank to borrow money at a given spread. Although theoretically the bank is adding a risk-free operation in its balance sheet (i.e.: a derivative contract and its associated replication strategy\(^{10}\)), the funding spread is not nil because the spread on the already issued debt \( D_1 \) is fixed and it cannot be updated. Burgard and Kjaer [3] theoretically justify a funding spread equal to zero in the replication strategy strategy by showing that, in a simplified balance sheet, the total spread paid after the start of the replication strategy would be such that it implies a risk-free rate for the incremental debt needed in it. A similar view has been proposed also in Nauta [20]. Nonetheless, this justification relies on the fact that all the outstanding debt is renewed at the moment of the replication’s inception, and in any case every time a change in the total outstanding amount occurs: in this way the overall lowering of the bank’s riskiness is redistributed on the old and new debt, so that the incremental funding spread would be zero (i.e.: the new debt marginally costs the risk-free rate). But, if the existing debt cannot be freely renewed, because it has its own expiry (in \( T \) in our setting) and the contract funding spread cannot be changed, then the marginal funding spread on the new debt will not be the risk-free rate.

Hull&White [13], [14] and [15] also think that the inclusion of funding costs is not justified: their argument hinges on the M&M theorem and, in the more recent works, on some example reproducing possible real cases of derivative pricing. We showed above that the funding costs do exist when some of the assumptions made by M&M are not matched in the real world, namely when considering a complex banking activity in a multi-period setting; as such funding costs have to considered, as it will be apparent in what follows. Besides, the authors seem to be more worried about the fact the funding can be seen also a benefit in the price of the derivative contracts: this may originate some arbitrage opportunities. We will show that we tend to agree with them as far as the inclusion of funding benefit is concerned, but we disagree with them on that funding cost should not be included as well. But more on this later on.

After having determined the funding costs that the creditors of the bank would rationally set on the debt issued by the bank to set up the replication strategy, let us see how this modifies, with respect to the framework outlined above, when we know that the bank needs to borrow money (oth-

\(^{10}\)We exclude here the model risk inherent to the replication strategy: the model can be actually a partially correct representation of the reality, so that the replication is not perfect. An example of model risk can be using the B&S model in a world where the underlying asset follows a dynamics with stochastic volatility instead of a deterministic one.
where $FV$ is defined as before with $\beta^F$ modified as:

$$\beta^F = \frac{uV_u - dV_d}{(u-d)B(1+r + f_21_{(\beta^F<0)})}$$

**FVA**, for the same considerations above, is still a positive number. This means that the short replication cost entails a lower value for the long position in the contract for the replicator. We can summarise these results and write the value to the replicator for a long/short position in the contract as:

$$\phi V^F(t) = \phi\frac{1}{1+r} [pV_u + (1-p)V_d] - \text{FVA}$$

$$= \phi V(t) - \text{FVA}$$

where $\phi = 1$ if the replicator is long the contract and $\phi = -1$ if it is short. Equation (47) implies a buy/sell price that the replicator would quote (even in absence of bid/ask spreads), since the value when going long the contract is lower then the theoretical risk-free price (equation (38)), while the value when going short is higher in absolute terms (i.e.: more negative) then the same theoretical risk-free price. In the Option Pricing Theory no-arbitrage bounds are identified typically by transaction costs, such as pure bid/ask spreads. We are here adding another component widening these bounds, which is given by the funding costs due to the default-riskiness of the replicator (see also Castagna [7], in the final remarks, and Castagna & Fede [10], chapter 10, for a discussion on this topic).

Let us now see how the contract is evaluated by the bank. Let $\Pi(t) = \phi[V^F(t) - \alpha S(t) - \beta^F B]$: if the bank correctly applies the replication strategy prescribed by the model, given that this is a perfect representation of the real world, then we should have that $\Pi(s) = 0$ for any $s \in [t,T]$. We are supposing here that the bank is considering also the funding costs in the replication strategy. Nonetheless, taking into account also the arguments of those that tend to exclude the **FVA**, we want to see what happens when the contract is inserted in the bank’s balance sheet and ascertain whether the marginal funding cost is actually the risk-free rate. The eval-

\[ V_u = aSU + \beta^F B(1 + r + f_21_{(\beta^F<0)}) \tag{40} \]

and

\[ V_d = aSD + \beta^F B(1 + r + f_21_{(\beta^F<0)}) \tag{41} \]

The funding spread appears only if the bank need to borrow money ($\beta^F < 0$).

Equations (33) and (34) are a system that can be easily solved for quantities $\alpha$ and $\beta^F$, yielding:

$$\alpha = \Delta = \frac{V_u - V_d}{(u-d)S} \tag{42}$$

and

$$\beta^F = \frac{uV_d - dV_u}{(u-d)B(1+r + f_21_{(\beta^F<0)})} \tag{43}$$

The value at time $t$ of the contract is then:

$$V(t) = \Delta S + \beta^F B \tag{44}$$

$$= \frac{V_u - V_d}{(u-d)} + \frac{uV_d - dV_u}{(u-d)(1+r + f_21_{(\beta^F<0)})}$$

After some manipulations, it is possible to express (44) in terms of discounted expected value under the risk neutral measure:

$$V^F(t) = \frac{1}{1+r} [pV_u + (1-p)V_d] + \text{FVA} \tag{45}$$

where **FVA** the funding value adjustment, equal to:

$$\text{FVA} = B[\beta^F - \beta] \tag{46}$$

So the value of the contract including funding is equal to the otherwise identical contract valued in an economy where the counterparty risk is excluded (and funding costs are nil) plus the funding value adjustment. When $\beta^F > 0$, then $\beta^F = \beta$ and **FVA** = 0, otherwise $|\beta| > |\beta^F|$ (since $f_2$ is a positive quantity) and the **FVA** is a positive quantity as well. This means that a bank that wishes to replicate a long position after going short in the contract, its replication cost is higher than the cost paid in a perfect market where default risk is excluded and $f_2 = 0$ always.

It is worth noting that if the bank wishes to replicate as short position in the contract, after entering a long position with the counterparty, then the contract is worth, in absolute value:

$$|V(t)| = \left| \frac{1}{1+r} [pV_u + (1-p)V_d] + \text{FVA} \right|$$

where **FVA** is defined as before with $\beta^F$ modified as:

$$\beta^F = \frac{uV_d - dV_u}{(u-d)B(1+r + f_21_{(\beta^F<0)})}$$
The valuation process relies on the following equation:

\[
\text{VB}(t) = E[X_1(ds_1 - f_1) + \Pi(T) + E] \\
= [X_1(ds_1 - f_1) + \Pi + E] (1 - PD_1)(1 - PD_V) \]

\[
= \sum_{j=1}^{L} \max \left[ X_1(\text{Rec}_{ij} - (1 + f_1)) + \Pi + E; 0 \right] \\
* p_j PD_1(1 - PD_V) \]
\]

\[
+ \sum_{j=1}^{L} \max \left[ X_1(ds_1 - f_1) + \text{Rec}_{v} EPE_V + ENE_V \right] \\
- \phi(\Delta S - \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) + E; 0 \] \]

\[\times p_j PD_1(1 - PD_1) \]

\[
+ \sum_{j=1}^{L} \sum_{i=1}^{L} \max \left[ X_1(\text{Rec}_{ij} - (1 + f_1)) + \text{Rec}_{v} EPE_V \right] \\
+ ENE_V - \phi(\Delta S - \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) + E; 0 \] \]

\[* p_j p_v PD_V PD_1 = E \]

where \(\Pi = E[\Pi(T)]\), \(\Xi = E[S]\) and \(ENE_V = \phi V^*(T) = \phi[p_{V_a} 1_{\{\phi V_a > 0\}} + (1 - p) V_a 1_{\{\phi V_a < 0\}}]\) is the expected negative exposure.

The loss is computed only on the \(EPE_V\), otherwise it is nil: the bank is unaffected by the counterparty’s default when the value of the contract is negative and the expected value of the contract fully enters in the equation.

Let

\[
\tilde{\mathcal{R}}_1 = \sum_{j=1}^{L} \max \left[ X_1(\text{Rec}_{ij} - (1 + f_1)) + \Pi + E; 0 \right] p_j, \\
\tilde{\mathcal{R}}_2 = \sum_{i=1}^{L} \max \left[ X_1(ds_1 - f_1) + \text{Rec}_{v} EPE_V + ENE_V \right] \\
- \phi(\Delta S + \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) + E; 0 \] \]

\[p_v \] and

\[
\tilde{\mathcal{R}}_{1,2} = \sum_{i=1}^{L} \sum_{j=1}^{L} \max \left[ X_1(\text{Rec}_{ij} - (1 + f_1)) + \text{Rec}_{v} EPE_V \right] \\
+ ENE_V - \phi(\Delta S + \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) + E; 0 \] \]

\[p_v \] We rewrite (48) in a lighter notation:

\[
\text{VB}(t) = E[X_1(ds_1 - f_1) + \Pi(T) + E] \\
= \left[ X_1(ds_1 - f_1) + \Pi + E \right] (1 - PD_1)(1 - PD_V) \]

\[
+ \tilde{\mathcal{R}}_1 PD_1(1 - PD_V) + \tilde{\mathcal{R}}_2 PD_V(1 - PD_1) + \tilde{\mathcal{R}}_{1,2} PD_V PD_1 \]

\[= E \]

(49)

The probability of default of the bank once the derivative contract is closed, can be computed by considering all the cases when the bank’s value drops below zero, in which case the limited shareholders’ liability floors the value at zero. We have that:

\[
PD_B = \sum_{j=1}^{L} 1_{\{R_{1,j} = 0\}} p_j PD_1(1 - PD_V) \]

\[+ \sum_{j=1}^{L} 1_{\{R_{2,j} = 0\}} p_2 PD_V(1 - PD_1) \]

(50)

\[+ \sum_{j=1}^{L} \sum_{i=1}^{L} 1_{\{R_{1,i,j} = 0\}} p_1 p_2 PD_V PD_1 \]

where \(1_{\{\cdot\}}\) is the indicator function and \(R_{1,i}^*\) is the \(b\)-th addend in the summation in each \(R_{1,i}^*\).

The Default of the Counterparty Does Not Imply the Default of the Bank

If the notion of the derivative contract is is much smaller than the asset \(A_1\) already included in the assets of the bank’s balance sheets, then (analogously to what have been seen above for the asset \(A_2\)) we have that:

\[
\tilde{\mathcal{R}}_1^* \approx \tilde{\mathcal{R}}_{1,2}^* \approx \tilde{\mathcal{R}}_1 \\
\text{and} \]

\[
\tilde{\mathcal{R}}_2 \approx X_1(ds_1 - f_1) + \text{Rec}_{v} EPE_V + ENE_V \\
- \phi(\Delta S + \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) + E. \]

The notation is the same as above.

After substituting these values, equation (49) can be written as:

\[
\text{VB}(t) = E[X_1(ds_1 - f_1) + \Pi(T) + E] \\
= \left[ X_1(ds_1 - f_1) + \Pi + E \right] (1 - PD_1)(1 - PD_V) \]

\[+ \tilde{\mathcal{R}}_1 PD_1(1 - PD_V) + \tilde{\mathcal{R}}_2 PD_V(1 - PD_1) + \tilde{\mathcal{R}}_{1,2} PD_V PD_1 \]

\[= E \]

(51)

By repeating the same reasoning we have presented in the case of two assets, let \([(ms_1 - f_1)X_1 + E] = H\) so that equation (51) can be written

\[
\text{VB}(t) = \]

\[H[(1 - PD_1)(1 - PD_V) + PD_V(1 - PD_1)] \]

\[+ \tilde{\mathcal{R}}_1 PD_1 + \tilde{\mathcal{R}}_2(1 - PD_1)(1 - PD_V) + \text{Rec}_{v} EPE_V \\
+ ENE_V - \phi(\Delta S + \beta^F B(1 + f_2 1_{-\phi \beta^r < 0})) \] \]

\[* PD_V(1 - PD_1) = E \]

(52)

By formula (5) and the definition of \(ms_1\):

\[
[(ms_1 - f_1)X_1 + E](1 - PD_1) + \tilde{\mathcal{R}}_1 PD_1 = E \]

The first term on the RHS of the first line in (52) is (by equation (5) and the definition of \(ms_1\)):

\[
H[(1 - PD_1)(1 - PD_V) + PD_V(1 - PD_1)] + \tilde{\mathcal{R}}_1 PD_1 \]

\[= H(1 - PD_1) + \tilde{\mathcal{R}}_1 PD_1 = E \]

Spring 2014
So that:

\[ \text{VB}(t) = \Pi(1 - PD_V)(1 - PD_V) + \\
+ \left[ \text{Rec}_V \cdot \text{EPE}_V + \text{ENE}_V - \phi(\Delta S + \beta^f B(1 + f_2 1_{(\phi \beta^f < 0)})) \right] \\
\times PD_V(1 - PD_V) \\
= 0 \]

(53)
or:

\[ \text{VB}(t) = \left[ \Pi(1 - PD_V) + \left[ \text{Rec}_V \cdot \text{EPE}_V + \text{ENE}_V \\
- \phi(\Delta S + \beta^f B(1 + f_2 1_{(\phi \beta^f < 0)})) \right] \cdot PD_V \right](1 - PD_V) = 0 \]

(54)

Since \( \Pi = \varphi V - \phi \Delta S - \phi \beta^f B(1 + f_2 1_{(\phi \beta^f < 0)}) \), and \( \varphi V = \varphi V^+(T) + \varphi V^-(T) \), equation simplifies to:

\[ \text{VB}(t) = \left[ \Pi - \text{Lgd}_V \cdot \text{EPE}_V \cdot \text{PD}_V \right](1 - PD_V) = 0 \]

(55)

where we have set

\[ - \text{Lgd}_V \cdot \text{EPE}_V \cdot \text{PD}_V = \text{CVA} \]

1 We also know that by definition of the replication strategy \( \Pi = 0 \) so that (55) holds only if we add an additional sum of cash at the inception such that we set the initial cost of the replication strategy at:

\[ \varphi V^+(t) \equiv \varphi V^-(t) - \text{CVA} = \varphi V(t) - \text{FVA} - \text{CVA} \]

(56)

As we have seen before for the replicator, the value of the contract depends on the side (long or short), on the funding costs (the FVA) and on the expected loss given the default of the counterparty (the CVA). Both quantities operate in the same way: they lower the value to the bank when it is long the contract, and they increase the negative value (i.e.: make it more negative) when it is short the contract. The two \( V^{Bid} \) bid and \( V^{Ask} \) ask values at which the bank can fairly trade the contract, expressing them in absolute terms, are:

\[ V^{Bid} = V - \text{FVA} - \text{CVA} \leq V \]

and

\[ V^{Ask} = V + \text{FVA} + \text{CVA} \geq V \]

which confirms the results in Castagna [7].

We can make the following considerations:

- in a multi-period setting, where existing debt cannot be renewed continuously, the production activity of derivative contracts by the bank, although theoretically risk-free,\footnote{We are aware of that the pricing is much more involved as the one we are presenting, since each component of the price is affected by the others and a simple decomposition as the one we are presenting here is possible only by disregarding them. Anyway, it is also true that the error we make in doing this is not substantial.} does not imply a marginal funding cost equal to the risk-free rate, but on the contrary a positive funding spread, if the replication strategy prescribes to borrow cash;

- when internally evaluating a derivative contract, if the banks pays for whatever reason a funding spread, then the standard replication theory that assumes a unique lending and borrowing risk-free rate has to be accommodated to include also a funding valuation adjustment (FVA);

- a credit valuation adjustment (CVA) enters also in determining the value to the bank;

- eventually the default of the bank does not matter. This means that the value to the bank is determined as it were not subject to default, even if it is taken into account when starting the evaluation process through \( PD_1 \) and \( PD_V \). As a consequence the debit value adjustment DVA is not present in the evaluation equation (56);

- the DVA is the CVA seen from the counterparty’s point of view. When, in the bargaining process, the bank has to yield the DVA as a compensation to the counterparty, it represents a cost and it cannot be compensated by the FVA, as it is usually affirmed nowadays in many works. On the other end, the FVA and the DVA are not two ways to call the same quantity, if not in the particular case of a loan contract (see Castagna [8] for a discussion on this point).

- there is no double counting of DVA and FVA: they operate with different signs on the value of the contract. The bank should try not to pay the first one, since it cannot be replicated under realistic assumptions;\footnote{Again, disregarding possible model risks.} the second one can be included in the value of the contract and should be considered as a production cost.

**Proposition 6.** When evaluating a derivative contract whose counterparty’s default does not imply the bank’s default, the correct and theoretically consistent value is given by the otherwise risk-free theoretically fair price
plus the credit and funding valuation adjustments. The sign of the adjustments depend on the long or short position taken into the contract by the bank. The bank’s default does not enter in the evaluation process, so that no debit valuation adjustment is considered.

The Default of the Counterparty Implies the Default of the Bank

Assume now that derivative contract has a notional such that the default of the counterparty implies the default of the bank. The valuation cannot be operated by pretending that the contract is a standalone operation without any regard to the remaining parts of the balance sheet. We need to resort to numerical procedures, but at least formally we come up with a valuation formula.

By rewriting equation (49), we have:

\[
VB(t) = E\{X_1(ms_1 - f_1) + \Pi(T) + E\} = [X_1(ms_1 - f_1) + \Pi + E](1 - PD_1)(1 - PD_V) + \epsilon = E
\]

(57)

where \(\epsilon = PD_1PD_V(1 - PD_V) + PD_1PD_V(1 - PD_1) + ENE(1 - PD_1)(1 - PD_V)\). Let \(PD_{1,2} = 1 - (1 - PD_1)(1 - PD_V)\). As before, we solve for the fair value of \(V\) and we get:

\[
\phi V^*(t) = \phi V(t) - FVA - EPD_{1,2} + [X_1(ms_1 - f_1)(1 - PD_1)(1 - PD_V) + \epsilon]
\]

(58)

\[
= \phi V(t) - FVA - EPD_{1,2} + db_V
\]

which is the value of the risk-free fair value plus the cash needed to compensate the other adjustments.

It is likely useful to calculate the debit value adjustment that the bank can be reasonably expected to be requested by the counterparty, to compensate the bank’s default risk it bears (the DVA from the bank’s point of view is the CVA from the counterparty’s point of view).

Assuming we are still in a market where perfect information is available to all agents, the expected recovery \(\text{Rec}_V\) on the derivative contract that the counterparty calculates, is:

\[
\text{Rec}_V = \sum_{j=1}^I \min \left[ \frac{X_1\text{Rec}_j + \phi(\Delta S_{1,\text{DF},j}) + \phi(\beta_1 B_{1,\phi,\rho,j}) + E}{X_1 + E} \right] p_{1,j}
\]

(59)

Equation (60) is worth a few comments: firstly, the derivative contract is excluded from the calculations, since it is considered for its expected liability aspect (the ENE), so that it cannot be part of the (possibly residual) value of the assets upon the bank’s default (this is why the recovery on the EPE never appears at the denominator). Secondly, if the bank went bust, that means that the issuer of the asset \(A_1\) defaulted, so that only a fraction of the face value \(X_1\) is recovered. Thirdly, the replicating portfolio can be an asset and/or a liability, depending on the signs of the quantity entering in it; this explains the indicator functions at the numerator of fraction which make the components of the replication portfolio an asset when their quantity is positive. If their quantity is negative, the absolute value adds to the total amount of the bank’s liability, which is at the denominator of the fraction. The total residual assets are divided by total liability to have the average recovery that each of them will collect (we are not considering priority rules in the distribution of residual assets).

The DVA is then:

\[
DVA = (1 - \text{Rec}_V) \times ENE \times PD_B = \text{Lgd}_V \times ENE \times PD_B
\]

(60)

which is the expected loss suffered by the counterparty on the expected exposure (ENE) when the bank defaults. The probability of the bank’s default depends on the probability of the counterparty’s default \(PD_V\): this is a circular relationship that causes no problem when the contract’s small compared to the total assets of the bank, so that we fall back in the first case examined. If the notional amount is big enough to produce the bank’s default when the counterparty goes bankrupt, \(PD_B\) should be adjusted by excluding the effect of \(PD_V\). When the counterparty’s default triggers the bank’s default, the value of the contract is positive to the bank, since it suffers a loss of the EPE such that the total assets are not able to cover the outstanding liabilities. But this means that the cases when the bank defaults after the counterparty’s default can never be applied to the ENE.

All this discussion on the DVA is to show that the claim usually stated in theoretical works (and, alas, often also in practice) are not very soundly grounded: the DVA is always a very difficult quantity to deal with and it may cause many inconsistencies when taken into account not in a proper fashion. As a general rule, the DVA is simply a cost for the bank corresponding to the counterparty’s CVA. The DVA can be computed by the usual tools only when the amount of the deal is such that the counterparty’s default does not trigger the bank’s default. In any case we would like to stress the fact that DVA is quite a different quantity from the quantity \(db_V\) defined before.

The following considerations are in order:

- when the default of the counterparty can trig-
The default of the bank, then the evaluation of the derivative contracts must consider the entire balance sheet of the bank: the process is quite cumbersome;

- in very general terms, the value of the contract is made of the components:

1. the theoretical value of an otherwise identical risk-free contract;

2. minus the FVA, computed as in the case examined before, depressing the value of a long position, or making more negative the value of a short position;

3. the maximum loss given the default of both the issuer of asset \(A_1\) and of the counterparty (and, hence, of the bank): this equals the equity capital \(E\) times the joint probability of default \(PD_{1,2}\). The maximum loss is limited to the equity due to the limited liability of the shareholders. This quantity can be roughly assimilated to the CVA component of the first case examined above and it operated in the same way on the initial value of the contract;

4. diversification benefit \(dv_f\), which accounts for the ability of asset \(A_1\) to contribute to the final value of the equity, if its issuer survives, and for the joint recovery when both the counterparty of the contract and the issuer default. In all cases the limited liability is considered, so that the total loss can never exceed the equity capital. This part can be roughly assimilated to the DVA, although some parts have to be more correctly assigned the the CVA, since \(e\) contains also cases when the loss suffered by the bank does not generate its default. Moreover, it is apparent that there is no reference to the loss given default suffered by the counterparty, because the evaluation is performed from the bank’s point of view and it is interested only in residual net values of the assets for itself, not in recovery values for the counterparty. In any case, the quantity \(dv_f\), as the DVA, increases the value of the contract when the bank has a long position in it, and reduces the negative value when it has a short position.

The justification adduced by some authors (see Hull&White [13], [14] and [15], for example) that the DVA is a benefit upon default due to the limited liability of the shareholders, is partially confirmed only in the cases when the derivative contract has a big impact on the balance sheet, otherwise the DVA should be considered as a cost (if paid). Also the identification of the FVA with the DVA (or a part of it) is not justifiable.

**Proposition 7.** When evaluating a derivative contract whose counterparty’s default can trigger the bank’s default, the correct and theoretically consistent value is given by the otherwise risk-free theoretically fair price plus the funding valuation adjustments, the maximum loss suffered on the joint default and the diversification benefit due to the possibility to cover losses with other non defaulting assets and the limited shareholders’ liability in case of bank’s default. The sign of the adjustments depend on the long or short position taken into the contract by the bank. The bank’s default enters in the evaluation process only as far as the maximum loss is considered. The diversification benefit, although playing a similar role, is only roughly ascribed to the DVA.

**Balance Sheet Shrinkage, Funding Benefit and No Arbitrage Prices**

Some authors (Burgard&Kjaer [2] and Morini&Prampolini [19], for example) recur to the “funding benefit” argument to justify the inclusion of the DVA into the valuation process of the derivative contracts by the bank. On the other hand, they see the DVA strictly related to the cash received when entering in the derivative contracts, so that DVA clashes with the \(-FVA\), in the sense that the former should be seen as a funding benefit.

For some types of derivative contracts, such as an uncollateralised short position in a plain vanilla put option, the quantity \(-FVA\) may be the same as the DVA. This happens when the underlying asset is repo-able, i.e.: can be bought via a repo transaction so that the actual financing cost is the implied repo rate, which is very near to the risk-free rate and the related cost of funding can be neglected. Hence, the only cost of funding refers to the cash the bank needs to pay the option’s premium (in this case it is zero, since the premium is received): in Castagna [8] we identified this part of the total FVA as \(FVA^P\), to distinguish it from the other component of the total funding value adjustment, \(FVA^U\), due to the financing costs of the position in the underlying, if any.\(^{14}\)

On the other hand, limiting the FVA to the only \(FVA^P\) component can be incorrect for all those contracts that can have a double sided value for either

---

\(^{14}\)Hull&White [13], [14] define these two quantities as DVA1 and DVA2, respectively.
counterparties and that generally start with zero value for both, so that no premium is exchanged between parties. In this case the FVA refers to possible future cash-flows and it definitely differs from the DVA.\textsuperscript{15}

We have seen above, confirming the results achieved in Castagna [5] and [7], that the DVA and the FVA operate with opposite signs on the value to the bank of the contract: in case the bank, when going long (short) the contract, is forced to pay the DVA, it makes the value higher (lower) than the level the bank would be willing to pay (receive) at the inception; at the same time, the acceptable value for the bank for a long (short) position is decreased (increased) by the FVA. If we accept the inclusion of \(-\text{FVA}\) into the valuation of derivative contracts, then there could be some overlapping of the DVA and the FVA, seen as a funding benefit and accounted for with the opposite sign.

It seems here that we have an opposite view on something upon which an agreement could (and should) be reached. The difference between our view and the one originated by the “funding benefit” argument lies in the balance sheet shrinkage: basically, this activity consists in the reduction of the outstanding debt as soon as a positive cash flow is received by the bank. In this way, both assets (the cash) and liabilities (debt) decrease and the bank pays less funding costs.\textsuperscript{16} We have shown in the two cited works that the “funding benefit” is a badly posed concept, and it would be better to talk about balance sheet shrinkage instead; moreover, the funding benefit can be used to replicate the DVA only under very specific (and very likely unrealistic) conditions.

When the valuation of a derivative contract relies on the balance sheet shrinkage policy, an immediate consequence is that any cash-flow, either positive or negative, is present valued by adding funding costs. The reason is simple: if the cash-flow is negative, the bank must finance it by borrowing money and pay at the expiry of the loan also the risk-free interests); if the cash-flow is positive, the bank can buy back some outstanding debt, thus partially saving the future expenditure of funding costs, which then add to the present value of the received cash.

The assumptions, implicit in the possibility to implement such balance sheet shrinkage policy, are that the bank has some outstanding debt to buy back, preferably with the same expiry as the derivative contract (a fairly realistic assumption) and that the buy back can be done relatively easily with low transaction costs (this could hardly be met in practice, since not all outstanding debt is liquid and dealing with tight bid/ask spreads in the market).

Assume the shrinkage policy is feasible and let us modify the evaluation process of contract by the bank by including it. In practice, if \(-\phi\beta^F > 0\) we modify equation (49) as follows:

\[
\text{VB}(t) = \mathbb{E} \left[ X_1 ms_1 - (X_1 - |\phi\beta^F Bi|)(1 + f_1) + \phi V(T) - \phi \Delta S + E \right] \\
= \left[ X_1 ms_1 - (X_1 - |\phi\beta^F Bi|)(1 + f_1) + \phi V(T) - \phi \Delta S + E \right] \\
\times (1 - PD_f)(1 - PD_v) + \mathbb{R}_1 PD_f (1 - PD_v) \\
+ \mathbb{R}_2 PD_v (1 - PD_f) + \mathbb{R}_{1,2} PD_f PD_v \\
= E 
\] (61)

where

\[
\mathbb{R}_1 = \sum_{j=1}^{l} \max \left[ X_1 \text{Rec}_{j} - (X_1 - |\phi\beta^F Bi|)(1 + f_1) + \phi V(T) - \phi \Delta S + E; 0; p_f \right],
\]
\[
\mathbb{R}_2 = \sum_{l=1}^{L} \max \left[ X_1 ms_1 - (X_1 - |\phi\beta^F Bi|)(1 + f_1) + \text{Rec}_{v} EPE_v + \text{ENE}_{v} - \phi \Delta S + E; 0; p_v \right],
\]
\[
\mathbb{R}_{1,2} = \sum_{l=1}^{L} \sum_{j=1}^{l} \max \left[ X_1 \text{Rec}_{j} - (X_1 - |\phi\beta^F Bi|)(1 + f_1) + \text{Rec}_{v} EPE_v + \text{ENE}_{v} - \phi \Delta S + E; 0; p_f p_v \right].
\]

In equation (61) the positive cash-flows related to the long bank account position \(-\phi\beta^F B > 0\), is deducted from the outstanding amount of debt \(D_1 = X_1\). This will generate a saving equal to the smaller interests paid on the reduced notional of debt. To simplify things, we did not analysed how the creditors of the bank would update the interest rate; moreover we did not consider the fact the the outstanding debt would probably deal in the market at a price lower from \(X_2\), which is the level we implicitly assumed.

In the standard replication argument presented above, \(\phi\beta^F > 0\) and the positive cash-flows are not invested in a risk-free bank account, but they are used to shrink the balance sheet, thus partially saving the future expenditure of funding costs, which then add to the present value of the received cash.\textsuperscript{15,16}

\textsuperscript{15}See Castagna & Fede [10], chapter 12, for the case of uncollaterlised swap contracts.

\textsuperscript{16}If the equity capital \(E\) is constant in time, balance sheet shrinkage is the same as the activity currently denoted as deleveraging. It should be noted, though, that the balance sheet deleverage can be alternatively achieved by only increasing the equity capital, thus simply decreasing the amount of debt capital as a percentage of the total liabilities.
we repeat the same reasoning as that one presented above, then it is easy to realise that the value of the contract to the bank will be:

$$\phi V^*(t) = \phi V^F(t) - CVA = \phi V(t) - \phi FVA - CVA \quad (62)$$

If the bank is long, the value will be lower, as in the previous case when the shrinkage policy is not adopted; when the bank is short, the value of the contract will be less negative to the bank, so that it will be ready to accept a lower premium from the counterparty to enter in the short position of the contract. In contracts where the value is always positive to one of the two parties, when the bank is long it will accept to pay the FVA but not the DVA; when it is short, it will yield the DVA to the counterparty, thus accepting a lower premium. But this is theoretically the same quantity as the (negative) of the FVA\textsuperscript{17} considering the balance sheet shrinkage policy, or the funding benefit as it is generally mentioned. The problem related to the double counting of the DVA and the FVA stems from this equivalence.

The balance sheet shrinkage policy seems an effective way for the bank to be aggressive in offering derivative contracts to the counterparties: it justifies the inclusion of the DVA, seen as a funding benefit, or $-FVA$. Although the two quantities may not fully coincide in any case, they are very similar in most of cases.

Nonetheless, this policy has many drawbacks that make it less enticing than it may appear at first look. We will briefly list some related problems:

- as mentioned above, it is not always possible to find suitable bonds (outstanding debt) to buy back in the market at fair prices;
- the policy produces counter-intuitive results, pushing the bank to progressively willingly lower the selling price of derivative contract when its creditworthiness worsens. It can be forced to do so because the counterparty can have enough bargaining power to ask for the inclusion of the DVA in the traded price, but this has to be seen as a cost and not an adjustment to lightheartedly accept;
- the shrinkage is possible only when there is a premium that must be paid by the counterparty (the asset can be financed in most of cases by a repo transaction, although this is not always feasible). In contracts such as uncollateralised forwards and swaps the bank still will pay funding costs related to the replication strategy and possibly also a DVA, but they act with opposite signs, as we have shown before;
- the funding benefit argument is basically a claim on the replicability of the DVA more than the justification of the inclusion of the FVA with reverse sign. The DVA can be replicated only under very restrictive conditions.

**Proposition 8.** If the balance sheet shrinkage policy is included in the evaluation a derivative contract, the correct and theoretically consistent value is given by the otherwise risk-free theoretically fair price plus the credit and funding (cost) valuation adjustments and the debit value adjustment (or alternatively the funding (benefit) value adjustment). The sign of the adjustments depend on the long or short position taken into the contract by the bank.

We already supported elsewhere\textsuperscript{18} the idea that the balance sheet shrinkage should not be considered a sound assumption when evaluating derivative contracts, which means that the DVA is treated always as a cost (if paid) and that the FVA never enters in the internal value of the contract with the negative sign (i.e.: as a funding benefit). This means that the internal value is the one we have defined in equation (56), when no balance sheet shrinkage was considered.

**EXAMPLE 4** We sketch here the effects of the assumptions made by the bank regarding the replication strategy (i.e.: feasible or not feasible balance sheet shrinkage) when a derivative contract is closed. To make things concrete, consider an option \( V \) on an asset \( S \), expiring in \( T \). Assume also that going short and long the underlying asset \( S \) can be done with repo/reverse-repo transactions, paying a repo rate (approximately) equal to the risk-free rate. In this case the funding is due only to the payment of the premium. The bank evaluates the “production” costs to sell it (i.e.: enter in a short position) and, applying equation (56), it gets:

- \( -V(0) = -10 \), value of the risk-free option evaluated with the chosen model (e.g.: B&S);
- \( -FVA = 0 \), since the premium is received;
- \( -CVA = 0 \), since no counterparty risk exists for a short position in an option;
- \( V^*(0) = -10 \)

\textsuperscript{17}In practice there are some differences between the FVA and the DVA, due to the fact the different recoveries are attached to the outstanding debt and the exposure of the derivative contract.

\textsuperscript{18}See Castagna [7].
The bank should sell the option at a price $P = 10$: in the is case the $P&L = P - V^* = 0$ at the inception. The counterparty of the bank in this contract is another bank with an equal bargaining power, so that it manages to charge DVA in the option price, since it wants to be remunerated for the counterparty risk related to the bank’s default probability (the DVA is the CVA from its standpoint). Given the bank’s default PD$_B$ and the loss given default of the option, the counterparty calculates the DVA = 1. So the traded price of the contract is:

- $V(0) = 10$
- $+DVA = 1$
- $-FVA = 0$
- $P = -10 + 1 = 9$

The bank sells the option at the price $P = 9$, so that the $P&L = P - V^* = -1$ at the inception, given the value the bank attached to this contract. Consider now the case when the bank assumes that it is able to operate the balance shrinkage as an ordinary and routine liquidity policy. If the bank thinks that the balance sheet policy can be hardly implemented in practice, the sell and buy prices at which it is willing to trade, are

$$V^{Bid} = 8.25 < V(0) = 10 < V^{Ask} = 10.$$  

If the bank evaluates contracts under the assumption that the of the balance sheet shrinkage is a feasible liquidity policy, regularly and effectively run in reality, then

$$V^{Bid} = 8.25 < V^{Ask} = 9 < V(0) = 10.$$  

In this case the bank is willing to sell the option at a price which is lower than the otherwise risk-free value, relying on the funding benefit it receives. This point may also raise some eye brows, besides the considerations on the feasibility of the balance sheet shrinkage policy.

**Conclusion**

We have proved that the classical results on the evaluation of investments by Modigliani&Miller [18] and Merton [16] hold only when the bank holds a single asset: in this case the value of the investment is independent from the capital mix (i.e.: the amount of debt and equity capital used to fund it). In a multi-period setting, for an additional investment the results do not hold (or they hold partially), so that the way it is financed does matter.
and the funding cost enter in the evaluation process. The main point in this framework is that the existing debt cannot be updated, until its expiry, to reflect the new risk of the bank’s total assets, so the total financing cost is higher than the level that would prevail if all the debt could be freely renewed. Moreover, in the real world the equity is used to cover losses generated indistinctly by all assets: the limited shareholders’ liability produces its effects on the evaluation process only if the held amount of the asset is large enough to trigger the bank’s default when the issuer goes bankrupt. This result gives a theoretical support to the well established practice in the financial industry to take into account funding costs and to charge them in the value of the asset when the bank’s bargaining power allows to do so.

For derivative contracts the results we derived are similar: the counterparty’s credit risk and the funding costs have to be included in the value of the contract to the bank, whereas the bank’s own default does not play any role, thus the debit value adjustment should be excluded. When the bank embeds the balance sheet shrinkage policy in the evaluation process, then the negative of the funding cost, meant to be a funding benefit, is included in the value, producing the same (often identical) effect of the inclusion of the DVA. Since in our framework the DVA never enters in the value, the possible inclusion of the funding benefit will never generate any double counting effects. Anyhow, we think that relying on the balance sheet policy is rather incautious, since we think that in reality it cannot be implemented on the base required by the derivative market-making activity.

It is maybe interesting to note that the conclusion we have reached is the same as in Hull&White [13], [14] and [15], although partially and for a different reason. In fact, the two authors exclude the FVA in any case, either it enters in the value as a cost or as a benefit, mainly relying on the classical investment evaluation theory’s result of Modigliani&Miller. Besides, they show in [15] the possibility of arbitrages when the funding benefit − FVA is accounted for in the value of the contract. In any case, their aim is to support the theoretical result of one price dealing in the market.

In our framework, the FVA increases the value of short positions (makes it more negative) and decreases the value of long positions. So the FVA is always a positive or nil value. When the balance sheet shrinking is considered as a feasible and ever implementable policy, then the FVA can also enter with a negative sign (as a funding benefit) thus decreasing the value of both short and long positions.

We believe that, although theoretically possible, the balance sheet policy should not be considered when evaluating derivative contracts. Firstly, the policy can be hard to be followed in practice and secondly it is not feasible when the contract is two sided (such as swaps) and it is dealt with no cash exchanged between parties. If and when it is actually implemented in practice, then a windfall for the bank occurs. If the buy-back of own bonds is not operated strictly following the rule, the bank should evaluate derivative contracts as we have indicated in formula (56), hence agreeing with H&W at least as far as the funding benefit is concerned. The funding benefit, if any, should be referred to the smoothing of the future cash-flows’ profile of the derivatives’ book, but this will be the object of future research.

ABOUT THE AUTHOR
Antonio Castagna is Senior Consultant and co-founder at Iason ltd and CEO at Iason Italia srl. Email address: antonio.castagna@iasonltd.com

ABOUT THE ARTICLE
Submitted: January 2014.
Accepted: February 2014.
References


chasing the future in Finance

Iason Ltd provides solutions that enable managers, risk-control officers and front office people to understand and value their portfolios. All solutions are built on the principles of soundness, accuracy and clearness.

PRICING

Modern Derivatives Pricing
Sound models to value derivative instruments according with the best and newest market practice (CVA, DVA and liquidity).

Complex Products
Sophisticated analytical models, supported by research, in order to effectively and efficiently evaluate all kind of exotic structures, accounting for all the relevant risk factors.

Revaluation of Books
The service is for those institutions not interested in building a trading and risk management system, but simply needing a fair revaluation of their (possibly very complex) derivatives positions.

RISK MEASUREMENT

Credit Risk
Innovative approach to Credit VaR (CreVar) modeling and computation.

Market Risk
Innovative and sophisticated solutions for market risk measure and management.

Counterparty Risk
Pioneers in the development of CCR and CVA measurement models and applications with devoted internal teams in many major banks.

ALM and Liquidity Risk
Consulting services to support ALM and treasury department of financial institutions. Innovative models taking into account the interrelation between funding, liquidity and CCR.

To learn more: www.iasonltd.com
To contact us: info@iasonltd.com. Follow us on:
@iason_ltd  company web page iason  iason.network
Special Interview

Alessandro Mauro
LITASCO, International Trading and Supply Company of Lukoil Group, Geneva
A talk with...
Alessandro Mauro

In this section Andrea Roncoroni is interviewing Alessandro Mauro, head of risk management for LITASCO, the International trading and supply company of the Lukoil Group, in Geneva. Starting from his personal experience Alessandro will provide a clear overview of energy risk management, denoting its differences with respect to financial risk management and its future developments. For a matter of clearness, R. and M. stands for Andrea Roncoroni and Alessandro Mauro respectively.

R.: Hi Alessandro, let’s start with your personal experience. Why did you get interested in risk management?
M.: Hi Andrea, thanks for your question. In the 90s second half risk management was an underdeveloped discipline, in general and in the energy sector specifically. To give you an idea the Global Association of Risk Professionals (GARP), now the leading professional organization in this discipline, was created in 1996. I felt this was a place where I could be creative and innovative. There were no books on the topic and it looked like a learn-by-doing sort of job, which I liked.

R.: How did you start?
M.: I was working in Scuola “Enrico Mattei” of ENI as a researcher in energy matters. In those years energy was becoming “financialized”, i.e. energy and finance were started having something in common. One day I found some articles about Value-at-Risk (“VaR”), at that time representing a new approach to risk measurement, and I started studying how VaR could be applied in the oil sector, and later to the energy sector in general. When I thought I had a decent understanding of the subject, I wrote an article. Those were the years of the VaR model created and marketed by Riskmetrics, a company founded by JP Morgan. It was a good product for financial institutions, but not for commodities producers and traders. It is enough to say that it was modelling exposure to energy commodities using only four US-based risk factors, where an oil trader will normally have exposures to dozens or even hundreds of risk factors. Hence, I started from risk measurement, which is only one of the three pillars of risk management.

R.: What are the other two pillars?
M.: Well they are risk identification and risk treatment. If you do not properly identify risks, you cannot measure them and you cannot “treat” them. Risk treatment is often understood as “hedging”, even if speculation is a form of risk treatment. At that time hedging in energy markets was taking its first steps, at least in Italy. ENI was starting, following advices coming from London-based banks, selling natural gas under innovative pricing formulae. At that point in time, even selling at fix price was to be considered innovative, after decades of oil-prices gas indexation. I was called on to help in performing analysis aimed at assessing how this new exposures could be hedged. In fact the project was called “hedging project”.

R.: Is energy risk management different from risk management in financial institutions?
M.: Yes, I am totally convinced about that. Risk identification, and partially risk measurement, are different. Risk identification is much more difficult, there is nothing as simple as ‘I am short 100 shares’. In fact exposure in energy markets does not depend uniquely on conscious choices made by traders. Many external actors, with their actions, have impact on exposure taking and value creation. You may say that even risk treatment using financial derivatives is a quite specialized one and
adapted to the energy markets peculiarity, due to the specific elements underpinning the underlying assets of those derivatives.

**R.:** What one can do to properly address the specificity of energy risk management?

**M.:** You need to have excellent resources, which means primarily excellent IT systems and human resources. As far as the latter are concerned, the choice is not simple. The talent pool is not vast and there is big competition in attracting talents, in a situation in which the request for energy risk analysts is increasing. The alternative is training your employees, either on the job or by selecting proper external training. In this respect, professional certifications in risk management are good choices.

**R.:** Which is currently the biggest challenge for energy risk management?

**M.:** I choose two. Having IT infrastructure and resources to support a complex business, understanding and mitigating other risks. Information Technology is crucial because nowadays simply you cannot run any business process without it. Risk management is a business process, but a very specialised one, and consequently it needs customized IT software. It is so specialized that the name of “ETRM” (Energy Trading and Risk Management) was coined. Being risk management a general business process, we need to make sure that we identify, measure and mitigate different risks, not just market risk. This part is often neglected, many people like to introduce more and more complexities in market risk models, and they overlook that at the same time they are increasing other typologies of risks (operational, fraud, etc).

**R.:** Thank you, Alessandro, for this interesting talk about energy risk management.

**M.:** Thank you too, Andrea.
in the previous issue

banking & finance

Fast Monte Carlo pricing of Nth-to-default swaps

Analytical Credit VaR Stress Tests

energy & commodity finance

Stochastic Optimization for the Pricing of Structured Contracts in Energy Markets

Pricing Spark Spread Option with Co-Dependent Threshold Dynamics

special interview to Marco Bianchetti

last issues are available at www.iasonltd.com and www.energisk.org