**RISK GOVERNANCE & CONTROL:
FINANCIAL MARKETS & INSTITUTIONS**

Postal Address:

Postal Box 36
Sumy 40014
Ukraine

Tel: +380-542-698125
Fax: +380-542-698125
e-mail: alex_kostyuk@virtusinterpress.org
www.virtusinterpress.org

Journal *Risk governance & control: financial markets & institutions* is published four times a year, in September-November, December-February, March-May and June-August, by Publishing House "Virtus Interpress", Kirova Str. 146/1, office 20, Sumy, 40021, Ukraine.

Information for subscribers: New orders requests should be addressed to the Editor by e-mail. See the section "Subscription details".

Back issues: Single issues are available from the Editor. Details, including prices, are available upon request.

Advertising: For details, please, contact the Editor of the journal.

*Risk governance & control:
financial markets & institutions*

ISSN 2077-429X (printed version)
2077-4303 (online version)

RISK GOVERNANCE & CONTROL: Financial markets and institutions

VOLUME 5, ISSUE 2, 2015, CONTINUED - 1

CONTENTS

AN EXTENSIVE EXPLORATION OF THEORIES OF FOREIGN DIRECT INVESTMENT

Patricia Lindelwa Makoni*

Abstract

The purpose of this study was to identify and examine the key foreign direct investment theories. The history and origins of FDI theories were considered, prior to dwelling in-depth on the theories themselves. FDI theories were classified under macroeconomic and microeconomic perspectives. Macroeconomic FDI theories emphasize country-specific factors, and are more aligned to trade and international economics, whereas microeconomic FDI theories are firm-specific, relate to ownership and internalisation benefits and lean towards an industrial economics, market imperfections bias. FDI theories are fairly complex to explain and apply. This paper is purely qualitative in nature, and attempted to explain the different FDI theories by providing an analyisis of the key theories used in many scholarly works**.

Keywords: Foreign Direct Investment (FDI), Product Life Cycle Theory (PLC), OLI, Investment Development Path Theory (IDP), Multinational Corporations (MNCs), Imperfect Markets, Eclectic Paradigm

**Department of Finance, Banking and Risk Management at the University of South Africa (UNISA), P.O. Box 392, UNISA, 0003, South Africa*
Tel: +27767538234
***This is article is derived from the author's PhD work-in-progress*

1 Introduction

According to the UNCTAD (2012), private capital flows consist of foreign direct investment (FDI), foreign portfolio investment (FPI), and other investment such as international banking flows and loans. An increase in international capital flows has resulted in faster financial globalisation than trade globalisation. As such, it has become more imperative to understand the underlying theories which help to explain this growth and movement in capital flows, mainly from the investor's perspective. Our focus will be on FDI as it has been the dominant capital flow, especially amongst developing countries.

This paper therefore presents a theoretical perspective of FDI. The first section gives an overview of FDI definitions. The second section discusses the historical background and the origins of FDI theories, while the third section gives a classification of FDI theories. The fourth section presents the macroeconomic FDI theories, followed by the microeconomic ones. The final section of this article gives a concluding summary to the study.

2 Definitions of Foreign Direct Investment

Foreign Direct Investment is defined as international investment made by one economy's resident entity, in the business operations of an entity resident in a different economy, with the intention of establishing a lasting interest (International Monetary Fund (IMF), 1993). According to the World Trade Organisation (1996), foreign direct investment (FDI) occurs when an investor based in one country (the home country) acquires an asset in another country (the host country) with the intent to manage that asset. The management dimension is what distinguishes FDI from portfolio investment in foreign stocks, bonds and other financial instruments. Alternatively, FDI can be considered as the ownership of 10 percent or more of the ordinary shares or voting stock of an enterprise which is usually considered to indicate 'significant influence' by an investor (IMF, 2000). This however differs from country to country and can even be determined by their policies, some of which restrict the levels of shareholdings of foreigners in local firms.

According to the World Bank (2004), Foreign Direct Investment is that foreign investment that establishes a lasting interest in or effective (active) management control over an enterprise. In its publication on The Benchmark Definition of FDI, the OECD (2008), defined FDI as the net inflows of investment undertaken to acquire a lasting management interest (10% or more of the voting stock) in a firm conducting business in any other economy but the investor's home country. Emphasis is also placed on the fact that the 10% threshold commonly referred to is recommended to ensure statistical consistency across countries. For

investment to qualify as FDI, emphasis is placed on the fact that the investor must meet the 10% voting share threshold commonly referred to, which as the recommended mainly to ensure statistical consistency across countries (UNCTAD, 2009). Lipsey, Feenstra, Hahn and Hatsopoulos (1999) had earlier commented that this "lasting interest" implies the existence of a long-term relationship between the direct investor and the firm, as well as a significant degree of influence on the management of the firm.

3 The history and origins of FDI theories

The origins of FDI are not fully understood. Although there are many schools of thoughts which have been used to explain this phenomenon, there is still no consensus on any superior or general theory of FDI.

FDI theory dates as far back as the early work of Smith (1776) [as cited in Smith, 1937] and Ricardo (1817), and was related to international specialization of production. In Smith's theory of absolute advantage, he explained that trade between two nations will occur if one country is able to produce and export goods using a given amount of capital and labour, more than its closest competitor (absolute advantage). However, Smith's theory did not explain how trade arose between countries where one country was not in the business of production. It is then that the work of Ricardo (1817) emerged, to explain FDI using the theory of comparative advantage. Ricardo was more interested in international factor movements as he was of the opinion that labour and capital were mobile domestically but not across borders. His theory was however flawed because it was based on the assumptions of two countries, two products and perfect factor mobility, but still did not justify international capital movements. This is therefore in direct contrast to the notion that, in a world typified by perfect competition, FDI would not exist anyway (Kindleberger, 1969). According to Denisia (2010), if markets were efficient, with no barriers to trade or competition; international trade would be the only mode of participation in the global markets. It is against this background that when Hymer (1976) published his 1960 thesis, he laid the foundation for other authors to come up with more plausible theories of FDI. In his arguments, he found that FDI was motivated by the need to reduce or eliminate international competition among firms, as well as Multi-National Corporations' (MNCs) wishes to increase their returns gained from using special advantages.

Mundell (1957) came up with a 2-sector model of international capital flows whereby capital flows were considered to be a substitute to international trade, resulting in factor price equalisation between countries. Mundell (1957) extended Ricardo's theory of comparative advantage by developing a model encompassing two countries, two products, two factors of productions and two identical production functions in both countries (Denisia, 2010). However, Mundell's model considered more short term, international portfolio type of investments rather than FDI, and therefore could not explain international production through FDI. Many of the earlier theories were based mainly on the U.S and Europe. To remedy the shortcomings of Mundell's model, Kojima and Ozawa (1984) contextualised their model in Japan, and advanced an argument that FDI occurs if a country has comparative disadvantage in producing one product, while international trade depends on comparative advantage.

The emergence and trend of post-Second World War investments (a shift from exporting to FDI) made by US firms to Western European countries between 1950 and 1970 can be explained using Vernon's (1966) product life cycle (PLC) theory. According to his theory, firms go through four production cycles: innovation, growth, maturity and decline. The underlying principles of this theory were technological innovation and market expansion; hence, while technology ensured the conceptualisation and development of a new product, the market size influenced the extent and type of international trade. In the initial stage, new products are invented, produced and sold in the internal markets. If the product is successful, production increases, new markets are penetrated and export develops. This is the transition from growth to maturity. It is also during this maturity phase that competitors emerge, and the product originator then sets up a production facility in the foreign market country to meet growing demand. Product standardisation occurs and incremental investment is then directed to any global site which offers the lowest input costs. After that, the product is exported back to the initial innovation country (exporter becomes importer as per the PLC) where it is eventually phased out, and the PLC starts all over again with the innovation of yet another product, since to emerge from the decline phase, the firm must be innovative again (Nayak & Choudhury, 2014). This is precisely what transpired when European firms began imitating the American products being exported to them; US firms had to set up production infrastructure in the local markets in order to maintain their market shares (Denisia, 2010).

Like other FDI theories, the PLC theory has its limitations. Primarily as pointed out by Boddewyn (1985), the product life cylcle is but just a theory because it was not tested empirically. The PLC theory also does not take into account all FDI determinants, in that it, for example, only explains the location aspects of manufacturing infrastructure but not their ownership (e.g. manufacturing under licence or set up subsidiaries). The theory is a simplified decision-making process, which assumes a smooth-sailing, sequential journey with no obstacles, and is more applicable to industries that use technology for its innovation (Buckley & Casson, 1976). The PLC

theory was further criticised for its failure to explain why it is profitable for a firm to pursue FDI rather than maintain its exporting strategy, nor the timing of the move to invest internationally (Nayak & Choudhury, 2014).

According to Boddewyn (1983), in the early 1980s, a cohort of researchers such as Casson (1979), Calvet (1981), Grosse (1985) and Rugman (1980) put forth their own versions of FDI theories. Although some of these researchers made a concerted effort to incorporate capital, location, industrial organization, growth of the firm, market failure, foreign exchange parity, investment portfolio and product lifecycle theories into one whole theory to attempt to explain the motives and patterns of FDI, most credit is given to Dunning's eclectic paradigm (theory) of international production (Boddewyn, 1983). The best-known theory of FDI is Dunning's 1977 Eclectic Paradigm in which he states that FDI occurs under different scenarios of ownership, locational and internalization advantages (OLI). This theory will be discussed in detail later, as it will be compared to more recent theories of FDI. It is for the above-discussed reasons that today, Popovici and Calin (2014) concluded that FDI theory is based on three integrative theories – the theory of international capital market, the firm theory and the theory of international trade. As such, it further necessitates the examining of FDI theories from two economic perspectives: the macroeconomic and the microeconomic views on FDI.

4 Classifying FDI theories

According to Denisia (2010), the macroeconomic perspective on FDI is that FDI itself is a type of cross-border capital flow, between home and host countries, and is captured in the balance of payments statement of countries, with the variable of interest being capital flows and stocks, revenues obtained from such investments. The microeconomic perspective on the other hand relates to the motives for investments across national boundaries, as seen from the investor's point of view. This follows on from Shin (1998) who critically reviewed existing theories of FDI and cited various scholars who classified FDI theories in a similar manner. Petrochilos (1983) classified macroeconomic FDI decisions based on variables which determine the investment decision (as cited in Shin, 1998, p.186), and mimic corporate investment behaviour, under the importance of the market size of the host country as measured by the GDP, growth of the market size, factor prices, interest rates, profitability and investor protection against tariffs and other such elements. According to him, the microeconomic determinants, drawn from the theory of industrial organisation (theory of the firm), are more concerned with firm and industry features which would give MNCs certain advantages over domestic firms. Caves (1971) gives examples of these features as including product differentiation, technology, the product life cycle and the size of the firm as measured by its sales or the value of its assets. Another scholar who classified FDI theories along the macro and micro economic views was Gray (1981). According to him, macroeconomic FDI theories emphasize country-specific factors, and are more aligned to trade and international economics, whereas microeconomic FDI theories are firm-specific, relate to ownership and internalisation benefits and lean towards an industrial economics, market imperfections bias.

5 Macroeconomic FDI theories

Lipsey (2004) describes the macroeconomic view as seeing FDI as a particular form of the flow of capital across national borders, from home countries to host countries, measured in balance-of-payments statistics. These flows give rise to a particular form of stocks of capital in host countries, namely the value of home-country investment in entities, typically corporations, controlled by a home-country owner, or in which a home-country owner holds a certain share of voting rights. Lipsey (2004) further explains that the variables of interest are the flow of financial capital, the value of the stock of capital that is accumulated by the investing firms, and the flows of income from the investments. Macro-level determinants that impact on a host country's ability to attract FDI include market size, economic growth rate, GDP, infrastructure, natural resources, institutional factors such as the political stability of the country, amongst others. The various theories are discussed below.

5.1 Capital Market Theory

This theory, also sometimes referred to as the "currency area theory", is considered one of the earliest theories which explained FDI. Based on the work of Aliber (1970; 1971), it postulated that foreign investment in general arose as a result of capital market imperfections. FDI specifically was the result of differences between source and host country currencies (Nayak & Choudhury, 2014). According to Aliber (1970; 1971), weaker currencies have a higher FDI-attraction ability and are better able to take advantage of differences in the market capitalisation rate, compared to stronger country currencies. Aliber (1970; 1971) further adds that source country MNCs based in hard currency areas can borrow at a lower interest rate than host country firms because portfolio investors overlook the foreign aspect of source country MNCs. This gives source country firms the borrowing advantage because they can access cheaper sources of capital for their overseas affiliates and subsidiaries than what local firms would access the same funds for.

While this capital market theory holds true in the case of developed countries such as the United States, United Kingdom and Canada, it was challenged by

later scholars on the basis of ignoring basic currency risk management fundamentals. A major criticism of Aliber's theory was made by Lall (1979) when he highlighted that the theory does not apply in the case of less developed countries with highly imperfect or non-existent capital markets, and those with heavily regulated foreign exchange rates. Also, Nayak and Choudhury (2014) allude to the fact that Aliber's theory does not explain investment between two developed countries with similar strength currencies, nor how developing country MNCs with weaker currencies are able to invest in developed countries with much stronger currencies. This they exemplified using the case of Chinese firms with sizeable investments in USA and the UK.

5.2 Location-based approach to FDI theories

Although FDI location is influenced by firm behaviour (a microeconomic element) insofar as the motives of its location, that is whether it is resource-seeking, market-seeking, efficiency-seeking or strategic asset seeking; the overarching decision is in fact taken on the basis of economic geography, which is a macroeconomic decision as it takes cognisance of country-level characteristics (Popovici & Calin, 2014). According to them, the theory explained the success of FDI among countries based on the national wealth of a country, such as its natural resources endowment, availability of labour, local market size, infrastructure and Government policy regarding these national resources. An off-shoot of this location-based theory is the gravity approach to FDI wherein it was assumed that FDI flows between two countries is highest, if those two countries are similar geographically, economically and culturally. Gravity variables such as size, level of development, distance, common language and additional institutional aspects such as shareholder protection and trade openness were regarded as important determinants of FDI flows (Popovici & Calin, 2014). This is however a very basic approach to the economics of FDI, because FDI flows are more complicated than just being about commonalities between nations. Being close together geographically may reduce transportation costs, but not necessarily the cost of labour, for example. Also, sharing the same culture may not necessarily result in increased profitability or trade between the two countries.

5.3 Institutional FDI Fitness theory

Developed by Wilhems and Witter (1998), the term FDI fitness focuses on a country's ability to attract, absorb and retain FDI. It is this country ability to adapt, or to fit to the internal and external expectations of its investors, which gives countries the upper-hand in harnessing FDI inflows. The theory itself attempts to explain the uneven distribution of FDI flows between countries. Wilhem's institutional FDI fitness theory rests on four fundamental pillars – Government, market, educational and socio-cultural fitness. At the base of the pyramid are socio-cultural factors which according to Wilhelms and Witter (1998), are the oldest and most complex of all institutions. Above that is education, which the authors affirm to being necessary in ensuring an attractive environment for FDI as educated human capital enhances R&D creativity and information processing ability. The actual level of education does not seem to matter much for FDI as the requirements are dependent on the various skills needs of projects to be undertaken. However what is certain is that basic education may impact on the productivity and efficiency of FDI operations, making formative education such as the ability to speak, hear, understand, interpret and implement instructions key for attracting FDI.

The third pillar, that of markets, accounts for the economic and financial aspects of institutional FDI fitness, in the form of machinery (physical capital) and credit (financial capital). Developed and well-functioning financial markets are hence a prominent feature in the MNC's investment decision-making process. The fourth and final pillar as put forth by Wilhelms is the Government. The role of a country's political strength plays the biggest role in the FDI game. Government fitness requires the adoption of protective regulation to manage market fitness. Popovici and Calin (2014) add that Government fitness is considered to include economic openness, a low degree of trade and exchange rate intervention, low corruption and greater transparency. If policies are hostile and unfavourable towards investors, MNCs will shy away from such countries as the political instability increases the risk burden on their investments. (Wilhelms & Witter, 1998). The authors concluded that although the pyramid is represented in a specific order, the four institutional pillars in fact are inter-related and interact in unison in different forms. For example, Government policies shape markets, education and sociocultural activities; market forces impact on the Government, education and socio-culture; education affects human capital and hence Government, markets and sociocultural norms and practices; and finally, sociocultural systems are the origin of Government, markets and education, respectively (Wilhelms & Witter, 1998).

Interestingly, the theory of institutional FDI fitness has been empirically tested mainly in the African context. Muthoga (2003) (as cited in Popovici & Calin, 2014), investigated FDI determinants in Kenya for the period 1967-1999, in their PhD thesis. The author found that economic openness, GDP growth rate, level of domestic investment, internal rate of return and availability of credit – all proponents of Government economic policies – enhance a country's attractiveness to foreign investors. Along the same ideologies, Musonera,

Nyamulinda and Karuranga (2010) evaluated the institutional FDI fitness model in the East African Community bloc, using Kenya, Tanzania and Uganda as their sample, and data drawn from 1995 to 2007. They found that for Tanzania and Uganda, FDI inflows were predetermined by more than a single country risk factor, such as population size, size of economy, financial market development, trade openness, infrastructure and other economic, financial and political risks. Their research also further refuted the perception that FDI inflows to Africa are attracted by natural resources. This was evidenced by that Tanzania and Uganda, both resource-poor countries, were also able to attract FDI on condition that their Governments fulfill two conditions: establish macroeconomic and political stability, and introduce an efficient regulatory framework, as well as eliminate corruption.

6 Microeconomic FDI theories

Lipsey (2004) also states that the microeconomic view examines FDI motivations from the investor's perspective, which would be similar to take a firm-level or industry-level perspective in making a decision. This micro-view thus examines the consequences to the investor, and to home and host countries, of the operations of the multinationals or of the affiliates created by these investments, rather than the size of the flows or the value of the investment stocks or investment position. These consequences arise from their trade, employment, production, and their flows and stocks of intellectual capital, measured by the capital flows and stocks in the balance of payments, although some proxies for the flow of intellectual capital are part of the current account (Lipsey, 2004). According to Das (n.d.), microeconomic FDI theories attempt to shed light on why MNCs choose to locate their subsidiaries where they do, and why they specifically seek to penetrate those locations. Many of these microeconomic FDI theories are all based on the existence of imperfect markets.

According to the firm-specific advantage theory, developed by Hymer (1976), the decision of an MNC to invest abroad rests on certain advantages at its disposal, such as access to raw material, economies of scale, access to labour, low transaction costs, intagible assets in the form of brands and patents, amongst others. It is in fact a firm-level (firm-specific) decision, rather than a capital market one (Das, n.d.). Hymer's theory which laid the foundation in explaining international production was also supported by scholars such as Kindleberger (1969) in his imperfect markets model; Knickerbocker's (1973) oligopolistic reaction theory of following the market leader; the internalisation theory of Buckley and Casson (1976) in an international context, as well as Dunning's (1974) eclectic paradigm. These theories are based on the same fundamental principle – the existence of imperfect markets, which then has a bearing on firm behaviour. As a result, other than Dunning's eclectic theory, no further attention will be given to them, as they are accounted for in Dunning's OLI paradigm.

6.1 The Eclectic Paradigm

This is probably the most well-known theory of FDI. On his way to winning the world acclaimed Nobel Prize, Dunning (1980) integrated various theories discussed above – being the international trade, imperfect markets (monopoly) and internalisation theories, and complemented these with the location theory, also briefly discussed earlier. According to Dunning (2001), in order for a firm to engage in foreign direct investment, it must simultaneously fulfill three conditions.

The firm should possess net ownership advantages over other firms serving particular markets. These ownership advantages are firm-specific and exclusive to that firm, in the form of both tangible and intangible assets such as trademarks, patents, information and technology, which would result in production cost reductions for the firm, enabling it to therefore compete with firms in a foreign country. These advantages were also emphasised by Hymer (1976) and Kindleberger (1969) in their market imperfections' theories on firm-specific and monopolistic advantages, respectively.

Secondly, it must be more profitable for the firm possessing these ownership advantages to use them for itself (internalisation), rather than to sell or lease them to foreign firms through licensing or management contracts (externalisation). Boddewyn (1985) refers to this as the internalisation condition. Finally, assuming that the preceding conditions are both met, it must be profitable for the firm to exploit these advantages through production, in collaboration with additional input factors such as natural resources and human capital, outside its home country; failing which, the foreign markets would then be served through exports, and local markets by domestic production. Location-specific factors have to be taken into consideration by the investing firms, as per the economic geography and institutional FDI fitness theories discussed under the macroeconomic FDI theories.

Boddewyn (1985) emphasises that the more a country's firms enjoy ownership advantages, the greater the incentive they have to internalise them, and the more profitable to exploit them outside their home country, then the higher the probability of engaging in FDI and international production. Because of the interrelatedness of the three conditions, it is important that they occur simultaneously, otherwise FDI cannot occur. The context and application of the Ownership, Location and Internalisation (OLI) paradigm differs from firm to firm, and hence the theory cannot be considered in isolation of theories which affirm the importance of the host country characteristics.

Although the Eclectic Theory was empirically tested by Dunning himself, it still has some limitations which critics have highlighted over the years. Boddewyn (1985) praised Dunning's theory for explaining the initial FDI decision by MNCs, but however laments the lack of explanation with regard to subsequent FDI increases, which may only require changes only in some but not necessarily all the OLI factors. In addition to this, Shin (1998) questions the applicability of the theory to LDCs which generally do not monopolistic firm-specific advantages such as high knowledge content. Another criticism of the eclectic theory is that it incorporates so many variables that it ceases to be operationally practical as it does not explain FDI at the firm, industry and country levels. This is on the basis that Dunning attempted to combine several complementary theories of market imperfection, which even on their own are already fairly complex (Nayak & Choudhury, 2014).

To address these shortcomings, Dunning (1981) then came up with the Investment Development Cycle or Path (IDP) theory, in which he proposed a link between a country's level of economic development and its investment positions. The IDP had four stages which followed a pattern similar to the product life cycle theory (introduction, growth, maturity and decline): no FDI; location-specific advantages arise due to Government intervention, hence attracting FDI inflows; domestic firms enjoy ownership advantages as wages rise, resulting in FDI outflows; countries finally become net outward investors in the fourth stage. The underlying hypothesis here is that due to the dynamic interaction between a country's GDP and its economic policies, these have the potential to affect both domestic and foreign firms' ownership advantages (Nayak & Choudhury, 2014). Despite these challenges, Dunning's eclectic theory however still remains the most recognised FDI theory.

Another criticism of Dunning's OLI paradigm was raised by Forssbaeck and Oxelheim (2008) when they questioned the menial role assigned to financial aspects in the FDI decision. In his defence, Dunning (1993) acknowledged the existence of a "financial asset advantage" which is a firm's knowledge of and access to foreign sources of capital, but points out that this merely a by-product of the size, efficiency and knowledge of MNCs, and not necessarily a standalone advantage. Forssbaeck and Oxelheim (2008) argue that a strong financial strategy enables a firm to minimise its cost and maximise availability of capital; thus by lowering the discount factor of any investment, that firm's likelihood to engage in FDI increases as a result of the financial advantage. To this end, they hypothesized that a firm will engage in FDI when, amongst other things, it has access to competitively priced equity, when it cross-lists its shares on a larger, more liquid stock market, when it enjoys strong investment credit ratings, and when it is able to negotiate reduced taxation and/ or attract subsidies. Forssbaeck and Oxelheim (2008) empirically tested their hypotheses using a sample of 1379 European non-financial firms' international acquisitions. In their series of tests, they evaluated what effect including finance-specific variables has on Dunning's OLI model, and found that there was a strong explanatory power of the financial variables, thereby concluding that financial factors are equally important in explaining FDI using the OLI model.

7 Conclusion

Having examined the available major FDI theories, it is clear that there is no single superior theory which comprehensively explains FDI. However, as it is necessary to conduct research from a specific theoretical background, it is hoped that the above classification and analysis of FDI theories provides an adequate grounding towards selecting the most appropriate theoretical framework for future scholarly work.

References

1. Agarwal, J.P. 1980, "Determinants of foreign direct investment: A survey", Weltwirtschaftliches Archiv, vol. 116, no. 4, pp. 739-773.
2. Aliber, R.Z. 1971, "The multinational enterprise in a multiple currency world", The multinational enterprise, pp. 49-56.
3. Aliber, R.Z. 1970, "A theory of direct foreign investment", The international corporation, pp. 17-34.
4. Boddewyn, J.J. 1985, "Theories of foreign direct investment and divestment: A classificatory note", Management International Review, pp. 57-65.
5. Boddewyn, J.J. 1983, "Foreign direct divestment theory: is it the reverse of FDI theory?", Review of World Economics, vol. 119, no. 2, pp. 345-355.
6. Buckley, P.J. & Casson, M. 1976, The future of the multinational enterprise, Macmillan London.
7. Casson, M. 1979, Alternatives to the multinational enterprise, Macmillan.
8. Caves, R.E. 1971, "International corporations: The industrial economics of foreign investment", Economica, pp. 1-27.
9. Das, B.K. n.d, , Theories of Foreign Direct Investment. Available: https://www.academia.edu/5898256/Theories_of_foreign_Direct_investment.
10. David, R. 1817, "Principles of Political Economy and Taxation", publicado en.
11. Denisia, V. 2010, "Foreign direct investment theories: An overview of the main FDI theories", European Journal of Interdisciplinary Studies, No. 3.
12. Dunning, J.H. 2001, "The eclectic (OLI) paradigm of international production: past, present and future", International journal of the economics of business, vol. 8, no. 2, pp. 173-190.
13. Dunning, J.H. 2000, "The eclectic paradigm as an envelope for economic and business theories of MNE activity", International business review, vol. 9, no. 2, pp. 163-190.
14. Dunning, J.H. 1998, "Location and the multinational enterprise: a neglected factor?", Journal of International Business Studies, pp. 45-66.
15. Dunning, J.H. 1993, "Internationalizing Porter's diamond", MIR: Management International Review, pp. 7-15.

16. Dunning, J.H. 1986, "The investment development cycle revisited", Weltwirtschaftliches Archiv, vol. 122, no. 4, pp. 667-676.
17. Dunning, J.H. 1981, "Explaining the international direct investment position of countries: towards a dynamic or developmental approach", Weltwirtschaftliches Archiv, vol. 117, no. 1, pp. 30-64.
18. Dunning, J.H. 1980, "Toward an eclectic theory of international production", The International Executive, vol. 22, no. 3, pp. 1-3.
19. Dunning, J.H. 1980, "Towards an eclectic theory of international production: some empirical tests", Journal of International Business Studies, vol. 11, no. 1, pp. 9-31.
20. Dunning, J.H. 1973, "The determinants of international production", Oxford economic papers, pp. 289-336.
21. Forssbæck, J. & Oxelheim, L. 2008, "Finance-specific factors as drivers of cross-border investment—An empirical investigation", International Business Review, vol. 17, no. 6, pp. 630-641.
22. Gray, H.P. 1981, Macroeconomic theories of foreign direct investment: An assessment, University of Reading, Department of Economics.
23. Grosse, R. 1985, "An imperfect competition theory of the MNE", Journal of International Business Studies, pp. 57-80.
24. Hosseini, H. 2005, "An economic theory of FDI: A behavioral economics and historical approach", The Journal of Socio-Economics, vol. 34, no. 4, pp. 528-541.
25. Hymer, S.H. 1976, The international operations of national firms: A study of direct foreign investment, MIT press Cambridge, MA.
26. IMF (International Monetary Fund), 2000, International Financial Statistics Yearbook. Washington, DC: International Monetary Fund, available from http://www.imf.org/external/data.htm
27. IMF (International Monetary Fund), 1993, Glossary of foreign direct investment terms, Washington, DC: International Monetary Fund, available from https://www.google.co.za/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&sqi=2&ved=0CCMQFjAB&url=http%3A%2F%2Fwww.imf.org%2Fexternal%2Fnp%2Fsta%2Fdi%2Fglossary.pdf&ei=261HVe36MaWa7gbDtYDQCg&authuser=1&usg=AFQjCNEey-x8eGj4tAJFCbwUzfyGHSejDg&sig2=bqE7Na9kLUOx2itoZ_1tSg&bvm=bv.92291466,d.d24&cad=rjt
28. Kindleberger, C.P. 1969, "American business abroad", The International Executive, vol. 11, no. 2, pp. 11-12.
29. Knickerbocker, F.T. 1973, "Oligopolistic reaction and multinational enterprise", The International Executive, vol. 15, no. 2, pp. 7-9.
30. Kojima, K. & Ozawa, T. 1984, "Micro-and Macro-Economic Models of Direct Foreign Investment: Toward a Synthesis", Hitotsubashi Journal of Economics, vol. 25, no. 1, pp. 1-20.
31. Lall, S. 1998, "The investment development path", Foreign direct investment and governments: Catalysts for economic restructuring, pp. 322.
32. Lall, S. 1979, "Multinationals and market structure in an open developing economy: The case of Malaysia", Weltwirtschaftliches Archiv, vol. 115, no. 2, pp. 325-350.
33. Lipsey, R.E. 2004, "Home-and host-country effects of foreign direct investment" in Challenges to globalization: Analyzing the economics University of Chicago Press, pp. 333-382.
34. Lipsey, R.E., Feenstra, R.C., Hahn, C.H. & Hatsopoulos, G.N. 1999, "The role of foreign direct investment in international capital flows" in International capital flows University of Chicago Press, pp. 307-362.
35. Mundell, R.A. 1957, "International trade and factor mobility", The American Economic Review, pp. 321-335.
36. Musonera, E., Nyamulinda, I.B. & Karuranga, G.E. 2010, "FDI Fitness in Sub-Saharan Africa-The Case of Eastern African Community (EAC)", Journal of International Business Research and Practice, vol. 4, pp. 1.
37. Nayak, D. & Choudhury, R.N. 2014, A selective review of foreign direct investment theories, (No. 143), ARTNeT Working Paper Series.
38. OECD, 2008, "OECD Benchmark Definition of Foreign Direct Investment", 4th ed., Organisation for Economic Cooperation and Development, pp. 234
39. Petrochilos, G.A. 1983, Foreign direct investment and the development process: The case of Greece, Gower Publishing Company.
40. Popovici, O.C. & Călin, A.C. 2014, "FDI theories. A location-based approach", Romanian Economic Journal, vol. 17, no. 53, pp. 3-24.
41. Rugman, A.M. 1980, "Internalization as a general theory of foreign direct investment: A re-appraisal of the literature", Review of World Economics, vol. 116, no. 2, pp. 365-379.
42. Shin, S. 1998, "A Critical Review of the FDI Theories", 아태연구, vol. 5, pp. 179-211.
43. Smith, A. 1937, The wealth of nations [1776], na.
44. Smith, A. 1776, "An inquiry into the nature and causes of the wealth of nations", London: George Routledge and Sons,
45. UNCTAD (2009), "UNCTAD Training Manual on Statistics for FDI and the Operations of TNCs – FDI Flows and Stocks", Vol 1, available from www.unctad.org/en/docs/diaeia20091_en.pdf
46. UNCTAD (2012), "Development and Globalisation: Facts and Figures", retrieved on 4 May 2015, available from http://dgff.unctad.org/chapter1/1.2.html
47. Vernon, R. 1966, "International investment and international trade in the product cycle", The quarterly journal of economics, , pp. 190-207.
48. Wilhelms, S.K. & Witter, M.S.D. 1998, Foreign direct investment and its determinants in emerging economies, United States Agency for International Development, Bureau for Africa, Office of Sustainable Development.
49. World Bank (2004), Development Education Program (Depweb): "Beyond Economic Growth Student Book", last accessed 7 March 2011, available from http://www.worldbank.org/depweb/english/beyond/global/glossary.html#24
50. World Trade Organisation (1996), Trade and Foreign Direct Investment, Press Release, Press 57, available from https://www.wto.org/english/news_e/pres96_e/pr057_e.htm

RISK AND/OR RESILIENCE MANAGEMENT

Jean-Paul Louisot*

Abstract

Risk management aims at managing all the uncertainties that may interfere with the objectives and missions of the organization. Resilience engineering aims at building its capacity to get over disturbances or stress while keeping the functionalities needed to survive, and possibly thrive. A recently open debate on an Internet blog launched by the risk managers of the Scottish Widows Bank seems to arise from what some professionals see as two competing branches of the management sciences. Whereas through the development of ERM – Enterprise-wide Risk Management – risk management is emerging at last to become a science, as well as an art and a practice, the mentioned above centered on the role of a newly forged name "resilience management". This opens a new front of the many debates that could derail the path to maturity of Risk Management as a science and reopen new silos much as Business Impact Analysis, BIA, or continuity management, might do if a clear distinction is not made between science, objectives and tools. However, because organizations are so interconnected today in the supply cloud that it is inevitable that they will face catastrophic risk and this is why resilience needs to be a core objective of any risk management plan? Whereas traditional risk management techniques alone may not be adequate to deal with such pervasive and insipient risk scenarios, resilience is ingrained into ERM.

Keywords: Resilience, Risk-Management, Enterprise-wide Risk Management, Strategic Redevelopment, Continuity, Risk to Reputation, Stakeholders

**Docteur ès Sciences de Gestion de la Sorbonne, Anc. Université Paris 1 Panthéon-Sorbonne, Institut Catholique de Lille, Directeur Pédagogique de CARM_Institute, Veuves*

1 Risk management objectives & resilience: how do they meet?

The core objective of any risk management effort is to ensure the organization survival whatever the circumstances it may be confronted to. In a financial approach to survival, one might say that all that will be needed is enough cash to go through the period following a damageable event; however this will be enough only if the organization can retain its stakeholders' trust and confidence through the episode. Although burning through cash may also be seen by stakeholders as a reduction in their value or if insurance was available at a nominal cost, an inappropriate use of assets.

However, even this prime objective may prove beyond the reach of some organizations under dire and exceptional circumstances; this is more specifically true when it comes to liabilities or environment damages. When referring to environment such events at the EXXON Valdez and more recently the explosion of the deep-sea petroleum-drilling rig Deepwater Horizon in the Gulf of Mexico come to mind. In fact, in most extreme catastrophes, the decision may be to limit their probability to a level such that stakeholders will "live with it." Their perception of the threat is such that the benefits of the activities of the organization prevail over the risk so that the organization retains what the British call the "social license to operate". With the explosion of social media, it has become of utmost importance for any organization to reach a proper balance taking into account all stakeholders' interests and expectations when making decisions at all levels; it may even prove the best way to ensure long term value for the stockholders.

On the other hand, a risk management objective limited to survival may well prove below the expectations of the stakeholders, even more so at a time when most large organization s recognize their exposure to procurement cloud collapse. When it comes to black swans, stakeholders will often recognize that a temporary disruption may prove unavoidable. But for exposures that could be qualified as moderate, that can be reasonably expected to occur over a five or ten years horizon, they remain volatile year in year out, the major economic players will request more and more to be satisfied that the tools of continuity put in place by their suppliers and sub-contractors will limit the consequences of any event on their own activities. Furthermore, the "survival objective" would not address the point of the economic result of the organization and the profits might take the stockholders on a rollercoaster that would not help the stock price for public companies. By the same token, economic partners might question

the long-term viability of the organization. In the case of non for profit, the rollercoaster ride by those in need can only exacerbate, for example, their need for government assistance which introduces more government debt and perhaps taxes. The sine wave magnifies with such disruptions. At the end of the day, it might discourage donors who would feel that the management of the organization *"does not know what they are doing."*

Other constituencies, like state, local authorities, and consumers could also be alerted by the perception of a chaotic short sighted management. In such a context, top management must assign other more forcing objectives to risk management such as maximum acceptable downtime, stabilizing financial results, corporate social responsibility, in other terms the impact of ethical choices or the organizational Values. In other terms, they must decide on the conditions they wish the organization to rebound after a serious, or even catastrophic. Depending on the organization, the goal may be to retain or gain market shares, maintain or improve profit, reach out to more people in need, etc.

With no need for long developments here, as top management set higher post-event objectives, risk management will require more resources, including finances. Therefore there in an increasing conflict between the general pre-event objective of economic efficiency and the choice of an improved rebound post-event. In other terms, the increase of overall "cost of risk" might be questioned by the owners and the financial analysts.

This is a clear illustration of the need to assess risk management efforts over the medium or long term, i.e. at least ten years, whereas the financial markets tend to force CEOs to manage short terms results, annual if not quarterly. If more resources are diverted towards risk management, these are not used to boost development and improve the return on capital and the short term results are less convincing.

To justify these « non-essential » efforts, it is necessary to broaden the horizon to take into account a major event could occur at any time and probably during a ten year period many major events could disrupt the organization. However, financial analysts should develop models taking into account the degree of potential failure when most firm are valued on a present value approach that has a built in hypothesis of an infinite future. What techniques can measure a long-term benefit to the company? To provide an answer to this question, it proved necessary to introduce a new concept that is gaining momentum in the strategic thinking at all levels; it is a way to assess a longer term management including elements such as sustainable development and corporate social responsibility. The concept of resilience is borrowed from metallurgy, but used also in psychology and sociology is resilience.

Any academic development on risk management, must refer to it, and a growing number of annual reports for global companies include it in their risk presentation. Although this explosion is relatively new in the last five years, as early as the nineties, the Canadian Auditors Associations provided the first definition in a guide for its members. In metallurgy, the resilience of a metal measures its capacity to regain its elastic qualities after a stress, mechanic or thermic; in social sciences, it measures the capacity of an individual, or an organization, to adapt to a rapidly changing environment.

Among the topics is the taunting issue that companies must improve flexibility without deviating from their core missions. However, the way in which they carry the mission may have to adapt to changing expectations. For example, it may not be efficient for a single product company diversify because that would dilute its efforts to provide optimal benefits to consumers for its desired product. In the same time, it should remain keenly aware should the consumers' taste move away from its one product and move ahead of its competitors or substitutes to anticipate such evolution.

As far as risk management is concerned, resilience measures or assesses the capacity of an organization to recoup after a major disturbance, or survive a crisis. This will typically require that the organization will be able to fulfil its major obligations to its main stakeholders, i.e.:

- society, comply with laws and regulations;
- personnel, retain employment levels & pays salaries;
- economic partner, secure contractual terms and conditions;
- stockholders, maintain profitability & dividends.

2 Evolution & explosion of Risk-management, what of change management?

To understand the role of the concept of resilience in the current risk management landscape, it is necessary to review briefly the evolution that risk-management went through during the last two decades. The explosion of risk management as a practice in many organization is such that many professional programs in universities now include risk-management as a fundamental branch of the management sciences. Few academics and professionals envisioned this evolution but clearly they have seen continuity and crisis management as becoming important processes in any organization. And is that not, what resilience is about?

In this context, risk management foremost objective is resilience. And this concept is applicable to all branches of industry and services, even though financial institutions seems at the forefront of the movement. One of the participants of a recent debate on the Internet suggest that that resilience is the surplus of the capacity of an organization to face a major disturbance, and the damages it might incur in a

given situation. The higher the difference, the margin, the higher the resilience. However, the limit of that vision is that is looks only at the financial resources, thus ignoring in fact human, technical, partners and information resources. Including, what will develop on the social media, and the impact on the organization's reputation.

Furthermore, the capacity of an organization to develop the right response in any event is directly linked with its suppleness and adaptability. This means that the management style and structure have a direct impact on the organisation resilience. A very hierarchical and bureaucratic organization whose personnel adheres strictly to fixed processes, with no leeway to act according to circumstances, even in the case of a crisis, where he must gain prior authorization from his hierarchy for any change like, for example, the implementation of a continuity plan. Only organizations that can offer a capacity for change can survive like a building erected to resist earthquake. This issue reflects a debate not yet resolved in the ISO 31000:2009 standard, where the need for "risk-owners", the operational managers to be in charge of managing the risks at their level, requires that they have both the responsibility and the authority it the ERM – Enterprise-wide Risk Management – is to be effective.

There is a new approach to resilience for which the concept goes far beyond risk-management, and even its thought process. In this context, resilience is the capacity of the organisation, and its staff, to adapt under all circumstances to changes, challenges, failures, and even ruptures or crisis. Thus resilience is key to the success of any strategy whatever the uncertainties of the future. However, resilience seems natural in some organisations' culture where individual initiatives are encouraged. May be an illustration will help at this stage, if one stumbles and falls, it might be worth catching the one Euro coin found on the ground before getting up! This is what Bertrand Robert has conceptualized under the construction of "creative rupture" that invites an organisation in such a situation to conduct a new SWOT analysis to reinvent its strategy so that it can take advantage of the evolution in its internal context (strength and weaknesses) and external context (threats and opportunities).

For example, in a positive/negative sense, the companies that supplied the infamous FEMA trailers had to change their ideas of distribution after the disaster from summer fun seekers to disaster homes. The issue of capitalizing on ruptures, is summarized in one question: Can the organization think of creative alternative uses of their products or creative new products that work within the framework of the economy or society after a disaster? With a daily regional paper in France, that would be destroyed with no chance of rebirth in case of a fire in its printing shop, we imagined to use the resources to develop an online paper and redevelop the land for apartments building as it was in a desirable location in a middle size city in the heart of France.

In the management of change, communication systems play an important role, all the more critical in the time of stress, it is essential that communication be a continuous process linking the organization with its stakeholders and building or comforting their trust, the communication on risk is only a part of a bigger picture. The communication in time of a crisis is efficient only so far as it rests on the institutional communication. Resilience requires also the development of a consistent and robust communication process with all stakeholders at all time.

3 The heart of the debate might it be differing visions of resilience?

In some industries, failure is not an option, as is the case for aeronautics and space, resilience is at the heart of any project, and the system must be developed to ensure success. In such a context, the entire organization must be governed, prepared, and trained for the resilience of all its operations. In such a situation, what is the relationship between resilience and risk management? The system is subjected to a system safety analysis that is part of the risk management framework.

If the risk-management program is part of the missions, developed during the conceptual stage and implemented even before the system is operational, then all risks should be identified and information provided even to those in charge of the conception; it is the most efficient way to reduce the probability of anything wrong happening and enhance the capacity of coping with the unexpected. On the other hand, if the implementation of risk management is delayed until the beginning of operations, it can identify emerging risks new threats, weaknesses, and exposures to prioritize them and propose an improvement process, but it may prove lacking as an add on rather than a built in process.

Some professionals and academics might suggest that in such a context, risk management may not be known under that name as it is part of the overall project engineering integrated into sound management incorporating "lean management", "legal compliance". As far as I am concerned what is described here is a real holistic and integrated risk management, i.e. ERM. In a proper system safety approach, risks are assessed, and treated as early as the conception phase and resilience a definite mission of the system. One could summarize in a proverb: "*It is better to prevent than to cure"…*

3.1 Resilience & standards

In the United Kingdom, the organization in charge of standardization, the BSI, recently published BS

65000[1] centered on resilience, but at the public level. There are a number of other standards that touch on the subject, without being exhaustive here[2]. From an attentive study of those documents, it springs that risk and resilience are related concepts; however, most professionals would agree that resilience is a key objective for all organizations as they tend to manage their resources so that they can adapt to changes, even radical changes and ruptures.

However, within this enlarged framework, resilience would extend far beyond risk-management to all the protective functions including, but not limited to, IT security, Physical security, health and safety, environment management, etc. Therefore, it must be envisioned within 360° approach combining culture, strategy, and change. It is in fact a definition of a mature ERM including continuity management, to reduce the impact of potential disruptions, and the learning process needed to make quick decision under stress even with scarce information in a time of a crisis.

3.2 Engineering resilience & ecology

The concept of resilience is so widely used that it covers different realities as the following:

- ***Engineering or reactive resilience***: Engineering resilience would be the velocity of return to a new stable state following a disturbance which implies to focalize on the speed of functioning. It is the issue of the organization reactivity, including the acceptable degraded functioning state and/or downtime a key parameter of any continuity management effort.
- ***Ecological or proactive resilience:*** Ecological resilience would take into account the unavoidable change within and without a system, like entropy, and aims at finding a new equilibrium within the new contextual framework thus reflecting the capacity of the organization to adapt and thrive. This proactive resilience rests on the flexibility of the organization that allows it to find positive answers to disturbances and could be also defined as a pre-event resilience founded on environmental sciences, thus key to sustainable development.

3.3 Risks & resilience

With this new paradigm over resilience, some professionals envision risk management as a part of resilience management. However such an approach would mean that risk management is limited to manage known risks. As a revision of Donald Rumsfeld remark, that would mean that risk management is expected to cope neither with the "known-unknown", i.e. emerging risks, nor with the "unknown- unknown", i.e. the Black Swans. Woods and Wreathall[3] develop a model or resilience in analogy with the stress-tension approach. They identify the initial response to an event as the uniform response of the whole organization when it has the capacity to meet the challenge. This is what they call the first level of reactive capacity, do they have in mind risk management? The second level of response is the capacity to adjust to a new situation and that would be the real deep resilience where the organization can no more rest on planned responses, processes, and resources whereas the answers go far beyond the limits of the first degree adaptation[4]. In their framework, risk management would only allow to anticipate on the probable and the possible, and true resilience would come only with the second order of change to adapt.

3.4 The right answer to different levels of disturbance

In a global and integrated approach, risk management cannot be limited in scope to known risks or to situations where prior continuity plans can be implemented to go over the difficulty encountered like an automatic pilot on a plane. In real life, it must address both reactive and proactive resilience. Efficient risk management applies to complex systems, interacting in an ever growing web of interdependencies, thus by essence in an unstable equilibrium. When the system moves away from its built-in balance, corrective measures must be implemented that will restore a balance. However, these swings out of the ordinary are still too often called "crisis" whereas all emergency situations are not conducive to crisis; naming crisis situations that are merely unexpected, or out of the probable, can have to negative impacts:

- A craze effect that could generate a crisis ; and
- A blasé attitude form staff that will not react as promptly as needed when a real crisis will loom.

This situation of unenlightened catastrophism » is clearly identified in a recent book by Dylan Evans[5] when he suggests that *"transforming low probability events in quasi-certainties when these events are perceived as particularly formidable by stakeholders is an approach of worst case scenario that can induce dreadful decisions."* This is the exact reason why it is crucial for the organizations survival and resilience that those in position of authority react un a gradual manner to the nature and potential severity of given circumstances.

Whereas the level of disturbance of a complex system is a continuum, we have chosen to illustrate the preceding remarks based on the description in four

[1] BS 65000 - Guidance for Organisational Resilience

[2] Business Continuity ISO 22301 , Risk Management ISO 31000, Crisis Management BS 11200 published Sept 2014, Resilience November 2014, and Business Collaboration BS 11000

[3] See bibliography

[4] Woods & Wreathall, 2008, p.146

[5] Evans Dylan, Risk Intelligence, New York (USA), Simon & Chuster, inc; (2012)

states proposes, among others, in the November 2007 issue of the Harvard Business Review, that is to say:

- ***Simple state***: It is the state for which the system has been set up, the nominal state and it is based on "best practices", an unstable equilibrium rarely maintained, but with the following characteristics:
  - ✓ Stability, clear cause/effect relationships;
  - ✓ Slow evolution, order and accomplishment;
  - ✓ Avoid complacency.
- ***Complicated state***: it is a state where expertise is essential and the domain of "good practices". It is the state in which operational managers, risk owners, can handle daily variations within the possible and it is characterized by:
  - ✓ Multiples possible responses, analysis of different solutions, readiness to listen to non-conventional thoughts;
  - ✓ ***Beware:*** making timely decisions is more important than to wait for the best.
- ***Complex state***: This is the state where innovating solution must be investigated ahead of the situation to plan for action, it is still in the hands of the operational managers but it requires a formal planning process. It is where business continuity plans are an efficient tool and it is characterized by:
  - ✓ Only experience feedback will lead to a good understanding the chain of events;
  - ✓ It is essential to size innovation, embrace creativity and new management models;
  - ✓ ***Beware:*** there is a risk to attempt a return to the pre-event situation without taking into account the new context.
- ***State of chaos or rupture*** : This is the state when acting fast is essential but with a strategic vision that is beyond operational managers and require the input of top management and even may be the board of directors, it is the level of disturbance that call for a *Strategic Redeployment Planning (SDP)* and the state is characterized by:
  - ✓ Impossibility to discern stable cause/effect pattern, no manageable schemes;
  - ✓ Some degree of order must be restored to return to a complex state, may be different from the pre-event equilibrium;
  - ✓ Communication must be transparent and specific with instruction coming from top management to implement swift strategic changes if need be *(it is not the time for dialogue)*
  - ✓ ***Beware:*** This is an ideal situation to implement innovations and strategic U turns *(change management)*

4 How to strengthen the resistance to risks? Ten "best practices"[6]

In terms of resilience, experience feedback is a key factor in the learning process. However, the experience of the organization may not prove enough to strengthen its resilience; lessons must be learned also from situations experienced by other organizations in the same industry, or similar contexts. At the end of the process, it is all about transmuting threats into opportunities to ensure optimal value creation.

However, it is important also to keep a clear mind and a vision far beyond the organization's backyard, best practices may be found in other branches, other part of the world, may be even from prospective substitutes. It is through a cooperation with the World Economic Forum that PricewaterhouseCoopers (PwC) has gathered a wealth of experiences from international experts and CEOs, mainly on corruption risks, cyber-risks, and risks in the procurement cloud, as well as natural catastrophes, but that could easily be extended to all other forms of risks. PwC came up with a list of ten "best practices" to strengthen the resilience of an organization:

- ***Educate permanently to instill the organization's Values***: Even if only a small part of an organization is going through a problem it can put in jeopardy the whole; this is why a strong common culture and shared values are vital for its resilience. For example, to limit risks linked to corruption the Royal HaskoningDHV[7] (RHDHV) has introduced a complete program of further education to embed integrity at all levels in the organization.

"Business integrity goes far beyond corruption, collusion and fraud; it includes also personal attitudes and behaviors"[8]

- ***Collaborate to promote information transparency:*** Within huge complex and global networks, transparent information from individual actors for decision makers enhances risk resilient decision making. The Barrick Gold Corporation[9] strengthened the effectiveness of its due diligence process regarding corruption and procurement through cooperation with NGO operating global networks; thanks to this decision they improved third parties' confidence in the information.

[6] Ed Simmons - Price Water Coopers - Access the World Economic Forum's full report here: Leading Practices Exchange: Managing Risk

[7] RHDHV is an international firm consulting in engineering and project management.

[8] Anti-corruption Practice 2 – UNDCP (United Nations Development Programme)

[9] Barrick Gold Corporation is a Canadian firm specialized in gold mining.

"A collective approach is recommended for transparency and due diligence to curb risks, and specifically those linked to corruption"[10]

• ***Promote zero tolerance for any breach on risks handling:*** There are some risky events that should be deemed unacceptable by organizations, and which all involved should clearly understand to be so. To limit corruption risks, Skanska[11]put in place a "five zero" internal policy. The five pillars to conduct a project are: zero non profitable projects, zero environmental hazard; zero workers comp, zero ethical breach, and zero quality defect.

"The three-step approach – pre-qualification, performance evaluation, and suppliers' development – allows to eliminate problematic suppliers." [12]

• ***Question permanently the hypotheses:*** At a time of accelerated change, the hypotheses on which were based the risk-resilient decisions are not valid anymore. The WHO questions continually the hypotheses that lead to the manufacturing of anti-flu vaccine anticipating virus mutations. In 2009, the aftermath of the H1N1 pandemic lead WHO to question the hypothesis according to which major pharmaceutical companies could increase their manufacturing in developing countries. Therefore, WHO helped the development of local productions of vaccine in these countries that resulted in an improved access to vaccination for the populations.

"It is essential to develop and implement a process to monitor and question the underlying hypotheses for the protection."[13]

• ***Support staff, so that they support the organization***: If the staff of an organization are not personally resilient, then the organization itself cannot be resilient, all the more during or in the aftermath of a crisis. During the nineties, following hurricane Andrew landing in Florida, local employers took many initiatives, as part of their pre-event planning, to help their local employees to recoup. These firms' assessment was that the assistance provided to their staff would help them support their employers and enhance the local economy resilience.

"It might be counter-intuitive that companies already stricken by a disaster increase their own burden by taking responsibility for additional risks and use its capital for the benefit of their employees. However, evidence proves that a swift return to work of the employees motivated for the restauration of their employer, it is probable that the employer would take longer to recover, and may be even never recover."[14]

• ***Make decisions on independent and reliable data bases and information:*** In a time of a rupture, or a crisis, data can be corrupted or lose credibility whereas robust informed decisions rest on the availability of specific and reliable data. The Japanese subsidiary of Deutsche Bank was able to make sound decision during the Fukushima Daiichi disaster thanks to independently collected data, at a time when other sources provided conflicting information.

"It is essential for organizations to rely on informed and exact assessment to make decisions in a time of a crisis."[15]

• ***Rehearse and prepare for a crisis with drills:*** Although organizations are not often challenged by crisis, they may still happen at any time and top management, as well as risk management professionals must be ready for action as early as the warning signs. There again, Deutsche Bank has developed a global program to train managers in crisis management through a series of real situation drills so that they develop the skills needed to act swiftly and with confidence in such situations.

"Specific plans are less useful than the capacity to develop a plan under stress when the situation requires. At a time when catastrophic risks are more and more global due to increasingly complex web of interconnections, a local and swift response is vital."[16]

• ***Set up alarm systems that will facilitate swift and early decision making:*** In most crisis, speed is of the essence. Therefore, information and systems that bring an early detection of threats are essential as they allow the organization to react even before the impact on its activity. To ensure a protection against an increasing number of cyberattacks, the US government has set up an agency specialized in the detection of emerging threats to alert critical infrastructure organizations, and the development of specific loss reduction strategies.

"Governments should set up agencies to alert economic agents on emerging cyber risks and loss control strategies for threats on critical infrastructures."[17]

• ***Place the responsibility for resilience on top management:*** Resilient organizations are able to identify trends, adapt to changing contexts, and ensure collaboration throughout the organization. But this can be achieved only if top management is involved in the project. Because of the ever increasing number of cyberattacks, among other threats, financial institutions have moved the responsibility of cybersecurity from the IT department to the CEO in all subsidiaries, with an overview by the board of directors. In the same time, the risk owners will have

[10] Anti-corruption Practice 4 – UNDCP (United Nations Development Programme)
[11] Skanska is a Swedish firm in the field of building infrastructures for International projects.
[12] Anti-corruption Practice 7 – UNDCP (United Nations Development Programme)
[13] Catastrophic Risk Practice 2 - United Nations Office for Disaster Risk Reduction (UNISDR)
[14] Catastrophic Risk Practice 4 - United Nations Office for Disaster Risk Reduction (UNISDR)
[15] Catastrophic Risk Practice 5 - United Nations Office for Disaster Risk Reduction (UNISDR)
[16] Catastrophic Risk Practice 15 - United Nations Office for Disaster Risk Reduction (UNISDR)
[17] Cyber Risk Practice 1 - United Nations Office for Disaster Risk Reduction (UNISDR)

to be in a position to react to risk that materialize at their level, hence the responsivity for managing risk within their scope of action must be coupled with the authority to act swiftly when something goes awry.

"Cyber resilient strategies should be developed at the board of directors' level in each organization so that it can efficiently identify trends, adapt continuously to business contexts, and be able to implement an efficient response to systemic chaos and ensure continuity of operations, to the best of its ability."[18]

- ***Share knowledge and experience within a network of reliable and trustworthy partners:*** Procurement clouds, critical interdependencies, systemic threats, these are only illustrations of the many networks the resilience of which is essential for the resilience of each organization which is only but a knot. Sharing information and know-how, sometimes very sensitive, in a network of partners with whom the work with trust and confidence, will improve the resilience not only of the organization, but also the entire network. The fight against cyber threats on critical infrastructures lead by the Australian government rests on a network of trusted and experienced organizations to share critical information and security strategies that facilitate quick responses and resilient defenses.

"Sharing knowledge with trust between private and public stakeholders improves the understanding of cyberattacks and the response necessary to curb those that could strike critical infrastructures."[19]

5 It is too early to risk a conclusion – a status report

The debate is still at an early stage and will be further fueled by the multiplication of sources and changing nature of uncertainties in the world imply that all organizations need to adapt permanently. However, it is all to cleat that the core mission of risk management will remain to assess all risks and offer solutions to curb their probability or their impact, not only known risks, but also emerging risks as well as those not even imagined today.

Events that occur rarely, or even that never occurred before, sometimes called Black Swans, are particularly difficult to handle and require a lot of attention; however, their management should not divert from the more common risks the systematic management of which not only reduces sources of inefficiencies in the organization, but also participate in the preparation of all staff for the handling of the dire situations. This will be effective only if the increasing independencies are recognized and the efforts are extended to and shared with all the partners as not a single organization is in a position to identify and mitigate all the threats that could disrupt the network emerging from within or without its known frontiers.

It is therefore necessary for the whole network to develop its adaptation capabilities based on a common foundation. But is there a way to achieve such a compatibility? It will require a social learning process engaging all citizens, an understanding of Paradoxal management[20], and a culture of co-evolution[21], rather than letting the Darwinian principle of natural selection be applied in these situations. It is essential to take advantage of punctual equilibrium offered by a wide functional diversity. It is then easy to understand how resilience puts an organization in a position to size the opportunities to improve or innovate that are offered by the changes in its internal and external context.

It is critical for an effective risk management that shared exposures and interdependencies are recognized and jointly managed. No organization lives in an autarkic state and the disruption suffered by some can impact the whole network like a shock wave. The British government has clearly identified the need when it ruled that risk managers in local authorities are requested to set up an active dialogue with all the risk-managers of public and private entities present on the territory. Le gouvernement britannique a bien identifié ce besoin en exigeant des risk-managers dans les territoires qu'ils mettent en place une concertation effective avec les risk-managers de tous les acteurs publics et privés actifs sur le territoire.

Developing a diagnostic or the entire system is all the more important that an organization may be surprised by a situation that may seem random or unpredictable whereas it is the inevitable consequence of a chain of events that was not identified proactively. True resilience, therefore, requires an in depth understanding of the internal and external contexts in which the organization operates, which is precisely one of the recommendation of the ISO 31000:20009 standard, and positioning the organization in these contexts. If this first step was to be ignored, the rest of the risk management efforts would be practically sterile, as the organization would not be in a position to curb its major threats, and enhance its best opportunities.

To reach the level of knowledge of the context needed, the approach must be dynamic and identify the main force lines of change. It is armed with this understanding of the systems potential weaknesses that the organization can identify its thresholds of fragility, assess how robust is the system, measure the

[18] Cyber Risk Practice 2 - United Nations Office for Disaster Risk Reduction (UNISDR)

[19] Cyber Risk Practice 5 - United Nations Office for Disaster Risk Reduction (UNISDR)

[20] Paradoxal management rests on the conviction expresses in a formula suggested by Bernard Nadoulek: "A company in which there is no order cannot survive, but a company without disorder cannot evolve." This conviction is the prerequisite in any collaborative management.

[21] Coevolution is the dynamic of evolution implying the interaction between genes and culture over a long time horizon.

degree of leeway at all levels. This in depth analysis allows to come to grips with anarchy, i.e. the stages of evolution that the organization must go through to adapt to the new realities, and take into account the various scales that influence the main stakeholders' interest. At the end, the most efficient approach would be to build the hyperspace of danger[22] of all the stakeholders to decipher the complexity of the network in which the organization is involved.

In this perspective, the traditional concept of resilience, called engineering here above, which would entail a return to the ex-ante situation, might prove dangerous if it means resisting to change, instead of adapting to it. Furthermore, in the first place, why return to the situation that let the dreaded event happen? The organization would not learn from its mistake and reinstate the even causes that generated the disturbance.

This is why assessing the level of disturbance as suggested above is essential for the best use of resources allocated to the management of risks. More precisely, the distinction between situations where prior continuity planning will return the organization to a new stability, without major change in the strategy, and those situations that call for a "strategic redeployment plan" where top management will be called upon to reassess the existing strategy.

In addition, contrary, to what some suggests, risk management cannot be limited to managing known risks, but also, and mainly, it must ready the organization to confront the unexpected, the unknown. The natural extension of traditional risk management, linked to insurance covers, is to work on Black Swans, and thus make the organization more robust.

Resilience i.e. the continuing development and adaptation of an organization in a context experiencing a constant evolution, has become a fundamental objective of top management. Risk-Management is the core function that contribute to building and preserving resilience provided it goes beyond the daily management of known risks and embraces a long term vision. However, this is possible only within the scope of ERM – Enterprise-wide Risk Management –, global and integrated extended to all the actors of the organization, internal as well as external, which supposes a share vision, shared values and a permanent effort of education and training.

There are many activities that contribute to the science of risk management, and its implementation within all organizations. They include continuity management, economic intelligence, health and safety, security, etc. that are all essential to the building and strengthening of resilience. However, risk management can be fully effective only if it is involved at all levels, strategic, tactic, and operational.

The board of directors must back all the efforts of risk management and tackle directly all the exposures whose impact is potentially strategic but they must be weary of only an inside-out vision and be sure to put in place instruments to obtain an outside-in vision that only questioning external and internal stakeholders can bring. It is the only way to take into account on all decision making the expectations and perceptions of stakeholders. It is the condition for the organization's strategy remain relevant and, provide products and services that meet the short, medium, and long term expectations of its economic partners while meeting its social responsibility obligations; thus keeping its "social license to operate".

Resilience finds its real existence, its soul, in the organization's reputation that must be patiently built and preserved throughout the tribulations of the world; but are not risk to reputation at the heart of risk management?

However important reputation is, a company might maintain its good reputation to the end but simply expire because of the weight of disaster? The case of Cantor Fitzgerald was called to my attention. In the WTC 9/11 disaster Cantor Fitzgerald lost almost all of its people. It was known for its great management and compassion—but in the end it suffered unmercifully from the disaster. It managed to survive, but how did it do so and is it the same compassionate company it was before?

References

1. Evans Dylan, *Risk Intelligence*, New York (USA), Simon & Chuster, inc; (2012)
2. Gauthier- Gaillard Sophie & Louisot Jean-Paul – Diagnostic des risques – AFNOR, 2nde Edition (2014)
3. Kauffman, S. A. (1991). Anti-chaos and adaptation. Scientific American, 265(2), 78-84.
4. Louisot Jean-Paul, 100 Questions pour comprendre la gestion des risques – AFNOR, 2nde Edition (2014)
5. Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social--ecological systems. Ecology and society, 9(2), 5.
6. Woods, D. D., & Wreathall, J. (2008). Stress-strain plots as a basis for assessing system resilience. Resilience Engineering: Remaining Sensitive to the Possibility of Failure, 1, 145-161.

[22] See Georges-Yves Kervern, "Latest advances in Cindyinics" – Economica Paris, 1994

THE ENHANCEMENT OF THE DETERIORATED SOUTH AFRICAN BOND OPTIONS MARKET

Coert Frederik Erasmus*

Abstract

Investments, especially those in the bond market, carry a level of risk. Risks in the bond market can be mitigated by transacting in option contracts. In the developing South African economy, trading activity of over-the-counter (OTC) bond options decreased significantly. Possible deteriorating factors and interventions to enhance OTC bond options were investigated in this research. Experts in the OTC bond option market were surveyed and the quantitative data collected was analysed with descriptive statistical methods. Results indicated that three factors were positively identified as deteriorating factors in the OTC bond option market and that five different interventions were possible to possible enhance this market.

Keywords: Bond Options, Deterioration, Developing Economy, Enhancement, Over-the-counter (OTC), South Africa

**University of South Africa (Unisa), AJH vd Walt Building, Office 5-104, Preller Street, Muckleneuk, Pretoria, 0003, South Africa*
Tel: +27 12 429 4949

1 Introduction

Investors entrust funds with assets manager that attempt to achieve profits as prescribed in individual fund mandates (Financial Services Board [FSB], 2007). The fund mandates prescribe the required rate of return and the financial instruments that asset managers are allowed to use as investment vehicles (FSB, 2007). The risk-return of a fund can be improved by adding financial instruments such as derivatives to a portfolio (Graf et al., 2014; Johnson, 2014). Option contracts were often used within a fund to reduce risk and increase the returns (Graf et al., 2014).

The standardised options contracts introduced by the Chicago Board of Options Exchange in 1973 became a popular instrument soon after introduction (Bodie et al., 2009). Almost more than a decade later, in 1987, standardised bond option contracts were introduced into the South African financial market (Bullard, 1987). The South African OTC bond option market flourished until 2001 (Johannesburg Stock Exchange [JSE], 2011).

However, the South African OTC bond option market deteriorated since 2001. In 2011 this market deteriorated to such an extent that it could be considered as insignificant (JSE, 2011).

This research attempts to determine the possible causes that led to the apparent deterioration of the South African OTC bond option market between 2001 and 2011. This research will further suggest possible methods that could again enhance this market to ascertain the benefits of an active OTC bond option market.

The benefits of bond options range from protection against economic crises to informational efficiency (Hakansson, 1999; Maniar, 2007). According to Almeida and Vicente (2009), investor preferences are communicated through options. Further, Cao and Ou-Yang (2009) found that without options, investors would not be able to take a position in the market based on informational events. Therefore it is possible that market activity could decrease if the possible benefits that bond options offered are foregone (Cao and Ou-Yang, 2009).

Several researchers recently focused on the pricing of bond options, all attempting to improve pricing models (Bermin, 2012; Becherer and Davis, 2010; Jamshidian, 2010; Milne, 2009). Only a few attempted to research the South African derivatives market (Bekale et al., 2015; Adelegan, 2009). This research is therefore unique. Firstly, because it identifies the determinants of the apparent deterioration of the South African OTC bond option market. And secondly, possible interventions are proposed to enhance the OTC bond option market.

The subsequent sections are structured as follows. Section 2 reviews the relevant literature and studies relating to the current research. Section 3 describes the research methodology used in, including information on the benefits and detriments. In section 4 the empirical results is presented and discussed. Finally, section 5 concludes the research.

2 Literature review

Derivative instruments, such as option contracts, are often used by financial institutions for arbitrage, speculation or hedging purposes in order to achieve their objectives in the financial market (Hull, 2008). The CEC (2009) found that financial instruments in the derivatives market have established risk management structures. However, even considering the established structures some stumbling blocks remain. Bezzina and Grima (2012) are of the opinion that derivatives are not fully utilised in the international financial markets. If derivative instruments are underutilised it could lead to decreased market activity.

2.1 Causes of the apparent deterioration of the South African OTC bond option market

The objective of the South African derivatives market is to improve liquidity and risk management in order to be more competitive internationally in the derivatives market (Adelegan, 2009). The OTC bond option market in South Africa however deteriorated to such an extent in the period between 2001 and 2011 that it was considered insignificant by 2011 (JSE, 2011). The different factors that could potentially improve the OTC bond option market are also the factors that could have led to the deterioration in this market. These factors include the underlying asset class of the derivatives, liquidity, pricing, exchanges and regulation.

Developed financial markets, such as the Australian Securities Exchange (ASX), have a flourishing OTC bond option market as well as an effective bond market (PIMCO, 2011). The example of the Australian market indicates that a relationship exists between an effective bond market and bond option market. Mu et al. (2013) reported in 2013 that South Africa had the largest bond market among the sub Saharan African countries.

The most liquid bonds in the South African bond market are the bonds issued by the South African government (Aling and Hassan, 2012). There are however fewer corporate bonds available (National Treasury, n.d.). The limited number of corporate bonds in South Africa negatively affects the liquidity in the market (Ball et al., 2011; Reid, 2009). The illiquidity in the South African corporate bond market therefore has a potentially negative effect on the South African bond option market.

Illiquidity in an underlying market often leads to inaccurate pricing in the derivative market (cf. Goyenko et al., 2011). In the South African context the lack of liquidity in the corporate bond market might have caused deterioration in the bond option market. Hull and White (2014) suggest that liquidity is essential to value a derivative accurately. Incorrectly priced derivatives are often the source of financial losses and incorrect financial information disseminated into the market (Dodd, 2009; Milne, 2009).

However, the illiquidity in the OTC bond option market could potentially be attributed to the nature of the OTC market as OTC markets tend to be illiquid (Choudhry, 2001; Deuskar et al., 2011; Stulz, 2005). The deterioration in the South African financial market in the period between 2001 and 2011 was therefore not necessarily the result of a lack of accurate pricing but perhaps a consequence of the nature of the OTC market. Deuskar et al., (2011) describe the nature of an OTC market to be illiquid. Further, Dodd (2012) and Stulz (2005) conclude that the investors in the OTC market are informed, indicating that investors have the knowledge to value derivatives in the OTC market accurately.

Ball et al. (2011) is however of opinion that the only difference between an OTC market and an ET market is the regulatory consequences of trading in the respective markets. In South Africa the OTC bond option market is standardised but unregulated (National Treasury, 2012). According to Adelegan (2009), South Africa does not have the capacity to regulate the financial market with prudential regulation.

Erasmus and Makina (2014) however stated that the conservative South African financial regulations limited the devastating effects of the 2007 subprime crisis since the regulation reduced the market exposure to foreign market risk. Botha and Makina (2011) reported that South Africa has two regulatory authorities, the South African Reserve Bank (SARB) and the Financial Services Board (FSB). South Africa also has self-regulatory authorities, namely the JSE and the central securities depository (CSD), Strate CSD. These authorities all form part of the innovations in the South African regulatory environment. The latest suggested regulatory framework is the Twin Peaks model that is focused on coordination, prudential regulation and market conduct (Goodspeed, 2013). Proper regulation has the ability to improve market efficiency (Caporale, Rault, Sova and Sova, 2015).

2.2 Enhancing the South African OTC bond option market

The primary objective of an investor who hedges a position is to limit the upside and downside risk of an investment (Hull, 2008). Investors who wish to hedge their positions often do not make use of the most suitable instrument (Dodd, 2009). If an investor uses an inappropriate derivative to hedge the risks associated with his or her open position, the hedge is often inefficient (Milne, 2009). To improve the marketability of OTC bond options risk management techniques should be re-evaluated, regulation should be improved, liquidity should be enhanced and

derivatives should be priced accurately (Adelegan, 2009).

Sherman (2009) concludes that the most effective method to limit market inefficiencies is effective regulation. Financial engineers however consistently engineer new methods to exploit regulation to their own benefit (Arestis and Karakitsos, 2009). The regulation to transact in the OTC bond option market must therefore be specific for the derivatives market in order to improve the market activity (Adelegan, 2009).

An increase in market activity could enhance OTC bond options in the financial market. Adelegan (2009) concludes that improved market activity would broaden the base of institutional investors and increase liquidity. Improved liquidity positively influences the price and pricing of a bond option. A liquid market allows reduced transaction costs and adaptable contract sizes (Deuskar et al., 2011). The improved liquidity, adaptable contract sizes and reduced transaction costs could allow more accurate pricing of OTC bond options and enhance this OTC bond options as hedging instrument in the South African financial market.

According to Amihud and Mendelson (1988) one of the most important components of a financial instrument is liquidity. Deuskar et al. (2011) concluded that a positive relationship exist in the bond option market between liquidity and asset prices. Lawton (2012) found that liquid assets can be transferred at a lower cost. Therefore, prices in the OTC bond option market would be reflected more accurately once the liquidity in the market improves.

The reviewed literature illustrated that a relationship exists between certain factors that influence the deterioration and the enhancement of OTC bond options. Factors such as liquidity, regulation and pricing could either have a beneficial or detrimental effect on the use of OTC bond options. Risk management techniques, the underlying asset and the type of option contract have an influence on the OTC bond option contracts. The next section discusses the methodology followed to determine the causes for the deterioration and possible interventions that could enhance the South African OTC bond option market.

3 Methodology

Empirical research was conducted in order to quantify the opinions of individuals who were employed by institutions that were active in the South African OTC bond option market at the time of the research. A quantitative research design is a design that allows the researcher to measure data to a definite degree of certainty (Crowther and Lancaster, 2009). Quantitative data is generally more exact compared to qualitative data as numerical data is described with standard statistical techniques (Crowther and Lancaster, 2009; Remenyi et al., 1998). The quantitative design also allows the collection of primary data. The primary data has the benefit of creating data to address a specific research question (McDaniel et al., 2008). A research instrument is required to collect primary data.

In the current research a survey was used as the research instrument. According to Remenyi et al. (1998), a survey is a quantitative tool that is used to collect data based on a respondent's knowledge, opinion or attitude. Surveys allow knowledge, opinions and attitudes to be reduced to numerical responses that are easily analysed by using statistics (Cooper and Schindler, 2008).

Data obtained from a survey is measured according to different scales. The ratio scale is the top level of measurement scales because it incorporates the abilities of all the lower-level measurements, namely the nominal, ordinal and interval scale and the ratio scale has a definite zero (Gershkoff, 2008). Additionally, a full range of statistical analysis techniques and mathematical tests is valid when data is measured with a ratio scale (Cooper and Schindler, 2008; Gill and Johnson, 2010; Remenyi et al., 1998).

The population for the current research was identified from data collected by the Profile Group (n.d.). The Profile Group regularly publishes a directory of the South African financial market employers. This directory includes all major financial institutions in South Africa. The population identified comprised all companies that were listed in the Profile Group Financial Market category of interest rate market members at the JSE.

The institutions that were involved in the option market were identified as issuers, market makers, intermediaries, participants and regulators. An issuer was defined as an institution that created a financial asset such as a bond (Wuite, 2009). Market makers bought and sold assets in the financial market at the quoted bid–offer prices (Van Heerden and Van der Westhuizen, 2008). Intermediaries in the financial market were purchasers of financial instruments on behalf of entities with surplus funds (Kalač et al., 2013). According to Adelegan (2009), participants acted in a similar capacity as an intermediary, with the exception that participants were retail investors or professional asset managers. However, participants did not buy on behalf of various entities but only for a specific entity (Adelegan, 2009). Lastly, regulators imposed legislation to combat market inefficiencies in the financial environment (Wuite, 2009).

The population of the current research comprised individuals who were employed by institutions who formed part of the South African OTC bond option market at the time of the research. Gill and Johnson (2010) state it would be unrealistic to distribute the questionnaires to the entire population; therefore, sampling is used. Since this research aimed to obtain data from experts in the field, who form part of a specialised population of OTC bond options investors, purposive sampling was the most appropriate (cf. Ishak and Bakar, 2014). The non-probability

purposive sampling technique is a superior non-probability sampling technique (Daniel, 2012).

The initial sample population consisted of 70 employers who employed individual whom were considered to be part of the interest rate market on the JSE. Each of these institutions was contacted to form part of the research. However, twenty-one of these institutions were only active in the interest rate market but did not participate in the interest rate derivatives market and were removed from the sample. The sample population was therefore reduced from 70 employers to only 49 employers. The sample population was distributed as indicated in Table 1.

Table 1. Sample population

Participants	26
Market makers	9
Intermediaries	9
Issuers	3
Regulators	2
Total Sample	**49**

Source: own composition

The questionnaire was distributed to individuals who were employed within this sample population. Questionnaires were sent to all the individuals who were identified from the sample population. Each respondent identified the extent of agreement with certain statements by indicating a percentage out of 100. The questionnaire sent to the individuals was constructed in order to obtain data on the causes of the deterioration in the South African OTC bond option market and the possible interventions that could enhance this market.

The questions that were identified as possible determinants of the deterioration of the South African OTC bond option market is listed in Table 2.

The possible factors that could enhance the South African OTC bond option market were identified and the determinants of the market enhancements are listed in Table 3.

The data collected with the research instrument was captured, stored and analysed (Bryman and Bell, 2011). The data collected during the current research was stored electronically in a password-protected Adobe Acrobat Professional X document. The data was then exported into Microsoft Office Excel 2010. Thereafter, the Microsoft Office Excel 2010 data was imported into the Statistical Package for Social Sciences (SPSS) program for data analysis. Lastly, the statistical results were interpreted in order to draw reliable conclusions, make recommendations and identify possible areas for further research. The next section reports the results from the statistically analysed data.

Table 2. Determinants of deterioration in the South African OTC bond option market

Question 1	Did bond issuers have a role in the deterioration of the OTC bond option market?
Question 2	Did the decline in the short-term interest rates since 2004 affected the OTC bond option activity?
Question 3	Was the OTC bond option pricing method that incorporated the volatility in the modified Black formula (1976) effective?
Question 4	Did wide bid–offer spreads have a deteriorating effect on the OTC bond option market?
Question 5	Was the OTC bond option market is efficiently regulated and overseen?
Question 6	Were International Swaps & Derivatives Association (ISDA) agreements sufficient to regulate the OTC bond option market?
Question 7	Did the ET bond option market crowd out the OTC bond option market?
Question 8	Were intermediaries active in the OTC bond option market?
Question 9	Did all institutional investors have access to the OTC bond option market?

Source: own composition

Table 3. Determinants of enhancements in the South African OTC bond option market

Question 10	Do investors fully understand the OTC bond option market?
Question 11	Is the OTC bond option market effectively integrated into the financial market?
Question 12	Would shorter-maturity OTC bond options stimulate the OTC bond option market?
Question 13	Is there regular interbank trading of OTC bond options?
Question 14	Did exotic options crowd out OTC bond options?
Question 15	Is the secondary bond market liquid and active?
Question 16	Is the OTC bond option market transparent?

Source: own composition

4 Results

The methodology used to collect data for this empirical research delivered comprehensive results as the respondents were representative of the South African OTC bond option market. The respondents from the sample population represented employers who acted as issuers (9%), intermediaries (9%), market makers (18%), regulators (9%) and participants (55%). The objectives of these

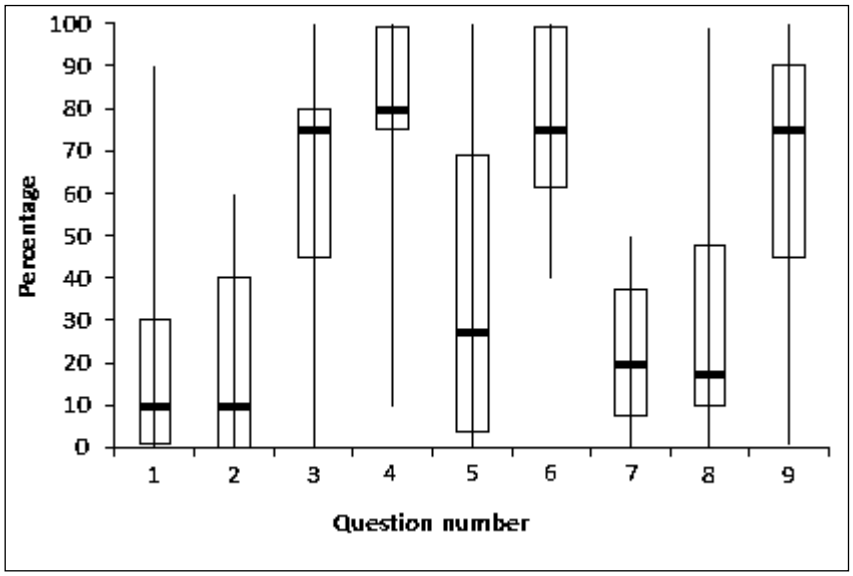
respondents who formed part of the market was speculative (37%), arbitrage (25%) and hedging (38%). The data obtained made it possible to identify the possible causes for the deterioration in the SA OTC bond option market and suggests the possible methods that could enhance this market.

4.1 Deterioration of the South African OTC bond option market

The literature review identified factors that could possibly have added to the deterioration of the South African OTC bond option market in the period 2001 to 2011. The factors that were identified were the lack of new bond issues, the declining short-term interest rate, the bond option pricing using the modified Black formula and incorporating the volatility skew, the effect that wide bid–offer spreads have on bond option activity, the market regulation and oversight, the standardised ISDA agreements used as binding contracts, the market in which OTC bond option contracts trade, the intermediary and broker participation in the market; and access to the market (Adelegan, 2009; Ball et al., 2011; Chinzara, 2010; Chordia et al., 2001; De Larosière, 2009; Durbin and Ng, 2005; Goyenko et al., 2011; Greenwood and Vayanos, 2010; Hearn and Piesse, 2012; IOSCO, 2010; JSE, 2011; Milne, 2009).

The deterioration was measured from the perspective of the respondent when answering question 1 to question 9 as presented in Table 2. Figure 1 is an illustration of the results obtained from the research instrument regarding the deterioration of the market.

Figure 1. Possible deteriorating factors in the South African OTC bond option market
(Box plot of question 1 to question 9)



Source: own composition

The respondents only perceived 33% of the determinants to be possible factors that contributed to the market inactivity. The survey results showed that unreasonable bid–offer spreads ($\overline{x_4}^{23} = 80.27$) resulted in the largest contributing factor to the deterioration in the market. The regulation and market oversight ($\overline{x_5} = 36.50$) was also identified as a factor that deteriorated the market as respondents did not deem the market to be efficiently regulated and overseen. The respondents found that the intermediaries also did not actively participate ($\overline{x_8} = 30.5$) in this market, which was also a perceived cause of deterioration.

The respondents did not agree that there was a lack of bonds issued ($\overline{x_1} = 22.36$) in the market to write underlying option contracts. Even though market regulation and oversight was perceived as inefficient the binding ISDA agreements ($\overline{x_6} = 76.4$) were considered sufficient. The declining interest rates ($\overline{x_2} = 16.55$) and bond option pricing method ($\overline{x_3} = 61.11$) were also not perceived as deteriorating factors. The results showed that the market was open to investors ($\overline{x_9} = 63.44$) and that exchange-traded trading alternatives ($\overline{x_7} = 22.22$) did not influence the market negatively.

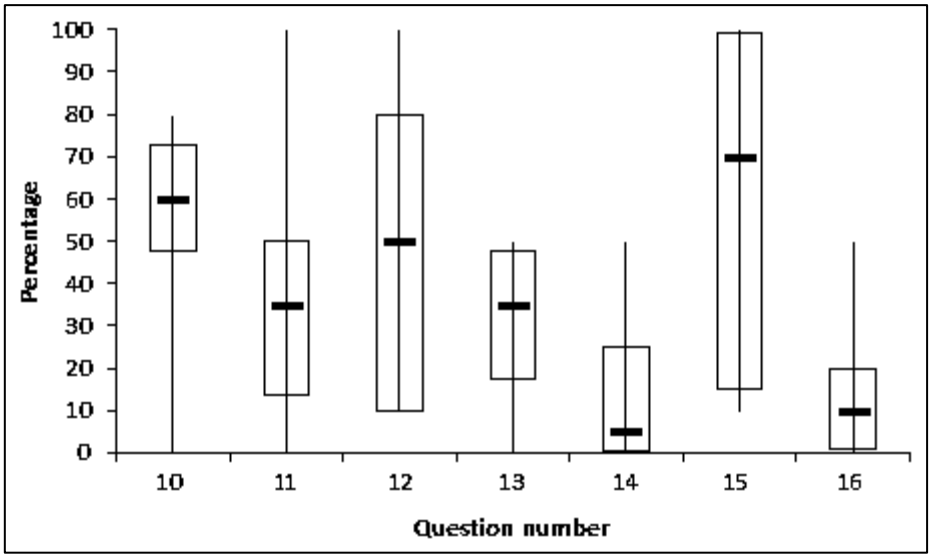
The effects of the determinants on the market deterioration were severe as the JSE (2011) stated the market could be considered as insignificant. The respondents were surveyed on the possible interventions that are possible to enhance this market.

[23] $\overline{x_q}$ is the average obtained from the results per question.

4.2 Enhancement of the South African OTC bond option market

There appears to be several possible methods available to restore the South African OTC bond option market. The literature suggested that investor education, market integration, shorter maturity options, liquidity and efficient regulation could possibly aid the enhancement of the South African OTC bond option market (Ameer et al., 2011; Arestis and Karakitsos, 2009; Bezzina and Grima, 2012; Choudhry, 2001; Cummins et al., 2001; Dodd, 2012; Dodd, 2009; Dodd and Griffith-Jones, 2007; IOSCO, 2010; PIMCO, 2011; Stulz, 2005). Figure 2 is an illustration of the results obtained regarding the enhancement of the market.

Figure 2. Factors perceived that could enhance the South African OTC bond option market
(Box plot of question 10 to question 16)



Source: own composition

According to the respondents, some methods could enhance the South African OTC bond option market. Results indicated that investors who were active in the financial market was uninformed ($\overline{x_{10}} = 55.0$) and did not fully comprehend the benefits of using OTC bond options, and therefore they did not make use of the instrument. The respondents were of the opinion that the market was non-transparent ($\overline{x_{16}} = 14.18$), which further decreased the use of this instrument. According to the respondents, at the time of the study, the OTC bond option market was neither fully integrated ($\overline{x_{11}} = 37.5$) nor was the instrument regularly traded between banks ($\overline{x_{13}} = 31.25$) in the financial market.

The secondary bond market was considered to be liquid and active ($\overline{x_{15}} = 60.82$) and would not currently have an impact on enhancing the market as the market was considered as functional. Respondents were however of the opinion that none of the interventions such as new exotic options ($\overline{x_{14}} = 14.0$) or the shorter maturity ($\overline{x_{12}} = 46.36$) OTC bond options would restore the OTC bond option market activity. Improved investor education, regulation, integration and liquidity could however potentially restore or improve activity in this market.

The empirical results of the current research were based on the opinions of the respondents who participated in this research. These respondents were experts in the field and active in the South African OTC bond option market. The respondents were employed to perform hedging, speculative or arbitrage transactions at the institutions where they were employed at the time of the research. The institutions formed part of the market as issuers, market makers, intermediaries, participants and regulators. The respondents provided their perspectives on the most probable causes for the deterioration of the OTC bond option market between 2001 and 2011, and possible methods for enhancing segments of this market.

5 Conclusion

The current research determined which factors led to the deterioration of the OTC bond option market in South Africa in the period between 2001 and 2011 and the interventions that could be effective to enhance the OTC bond option market. Table 4 is a summary of the actual factors that led to the OTC bond option market deterioration and the effective interventions that could enhance the OTC bond option market activity as perceived by the respondents from the sample population.

Table 4. Deteriorating factors and possible enhancement interventions

Actual factors	Effective interventions
Inactive intermediaries	Improved liquidity
Unreasonable bid–offer spreads	Investor education
Regulation and market oversight	Improved transparency
	Increased interbank trading
	Improved integration in the financial markets

Source: own composition

The OTC bond option market in South Africa could potentially be an effective tool for risk management (Adelegan, 2009). However, factors such as inactive intermediaries, wide bid-offer spreads as well as inefficient regulation and market oversight had a deteriorating effect on the market. Currently, the lack of liquidity and market integration in the South African OTC bond option market prohibit investors from benefiting from the use of OTC bond options. The limited market integration makes hedging inefficient to investors.

It is recommended that issuers, intermediaries, market makers, participants and regulators act on the following recommendations to restore the South African OTC bond option market:

- increase the liquidity in the OTC bond option market;
- encourage regular interbank trading of OTC bond options;
- encourage continuous education for investors and traders;
- improve the transparency in the OTC bond option market; and
- integrate the various financial instruments effectively in the financial markets.

The liquidity intervention could be addressed by encouraging intermediaries to actively participate in the market. Furthermore, liquidity would be enhanced if the wide or one-directional bid-offer quotes by primary dealers are discouraged. Liquidity would also be improved once regular interbank trading of the OTC bond options take place. The findings showed that shorter maturity OTC bond options would not enhance the market but it could promote interbank trading. Banks should also be encouraged to transact directly in the market as the use of brokers or intermediaries reduces transparency.

Transparency is reduced in some transactions as certain institutions make use of intermediaries. This is to limit the information about their financial position that is disseminated into the financial market. Institutions should be encouraged to transact in the market in their unique legal capacity. Once the OTC bond option market is perceived as liquid and transparent, OTC bond options could possible integrate into the financial market. The transparent market could become attractive to investors and a liquid market would extend itself to properly priced financial instruments that could stimulate the market.

Investors are however perceived not to be fully informed of the market functionalities. The South African OTC bond option has changed considerably since inception in the 1980s. It is therefore recommended that traders and investors who are active within the OTC bond option market form part of a professional body that encourages continuous professional development.

If the bond option market was restored, bond options could potentially be used to hedge, speculate or derive arbitrage profits. Though these recommendations may restore the South African OTC bond option market it is recommended that further research be conducted on regulatory changes and its effects on OTC derivative markets, the preferred derivative instrument for risk management, and identification of appropriate CPD interventions for South African OTC bond option market participants.

References

1. Adelegan, O.J. (2009), "The derivatives market in South Africa: Lessons for sub-Saharan African countries", International Monetary Fund, working paper series no. WP/09/196, Washington, DC.
2. Aling, P. and Hassan, S. (2012), "No-arbitrage one-factor models of the South African term structure of interest rates", South African Journal of Economics, Vol. 80 No. 3, pp. 301–318.
3. Almeida, C. and Vicente, J. (2009), "Are interest rate options important for the assessment of interest rate risk?", Journal of Banking & Finance, Vol. 33 No. 8, pp. 1376–1387.
4. Amihud, Y. and Mendelson, H. (1988), "Liquidity and asset prices: Financial management implications", Financial Management, Vol. 17 No. 1, pp. 5–15.
5. Arestis, P. and Karakitsos, E. (2009), "Subprime mortgage market and current financial crisis", Cambridge Centre for Economic and Public Policy, working paper series no. WP08-09, Cambridge.
6. Ball, R.T., Hail, L. and Vasvari, F.P. (2011), "Equity cross-listing in the US and the price of debt", European Corporate Governance Institute, working paper series no. 274/2010, Brussels, Belgium.
7. Becherer, D. and Davis, M. (2010), "Arrow-Debreu prices", Encyclopaedia of Quantitative Finance, Wiley, Enschede, Netherlands.
8. Bekale, A.N., Botha, E. and Vermeulen, J. (2015), "Institutionalisation of derivatives trading and economic growth: evidence from South Africa", Economic Research South Africa, working paper series no. 505, Cape Town.

9. Bezzina, F.H. and Grima, S. (2012), "Exploring factors affecting the proper use of derivatives: An empirical study with active users and controllers of derivatives", Managerial Finance, Vol. 38 No. 4, pp. 414–435.
10. Black, F. (1976), "The pricing of commodity contracts", Journal of Financial Economics, Vol. 3 No. 1, pp. 167–179.
11. Bodie, Z., Kane, A. and Marcus, A.J. (2009), Investments, 8th edn., McGraw Hill/Irwin, New York, NY.
12. Botha, E. and Makina, D. (2011), "Financial regulation and supervision: Theory and practice in South Africa", International Business and Economics Research Journal, Vol. 10 No. 11, pp. 27–36.
13. Bryman, A. and Bell, E. (2011), Business research methods, 3rd edn., Oxford University Press, New York, NY.
14. Bullard, D. (1987), "Investment basics XXI Options trading in the gilt-market", The Investment Analysts Journal, Vol. November, pp. 40–42.
15. Cao, H.H. and Ou-Yang, H. (2009), "Differences of opinion of public information and speculative trading in stocks and options", Review of Financial Studies, Vol. 22 No. 1, pp. 299–335.
16. Caporale, G.M., Rault, C., Sova, A.D. and Sova, R. (2014), "Financial development and economic growth: Evidence from 10 new European Union memebers", International Journal of Finance & Economics, Vol. 20 No. 1, pp. 48–60.
17. Chinzara, Z. (2010), "Macroeconomic uncertainty and emerging market stock market volatility: The case for South Africa", Economic Research Southern Africa, working paper series no. 187, Grahamstown.
18. Chordia, T., Roll, R. and Subrahmanyam, A. (2001), "Market liquidity and trading activity", The Journal of Finance, Vol. 56 No. 2, pp. 501–530.
19. Choudhry, M. (2001), The bond & money markets: Strategy, trading, analysis, Butterworth-Heinemann, Oxford, UK.
20. Cooper, D.R. and Schindler, P.S. (2008), Business research methods, 10th edn., McGraw-Hill/Irwin, New York, NY.
21. Cummins, J.D., Phillips, R.D. and Smith, S.D. (2001), "Derivatives and corporate risk management: Participation and volume decisions in the insurance industry", The Journal of Risk and Insurance, Vol. 68 No. 1, pp. 51–91.
22. Daniel, J.N. (2012), Sampling essentials: Practical guidelines for making sampling choices, Sage, Thousand Oaks, CA.
23. De Larosière, J. (2009), "The high-level group on financial supervision in the EU", in European Commission, viewed 28 February 2013, from http://ec.europa.eu/internal_market /finances/docs/de_larosiere_report_en.pdf
24. Deuskar, P., Gupta, A. and Subrahmanyam, M.G. (2011), "Liquidity effect in OTC options markets: Premium or discount?", Journal of Financial Markets, Vol. 14 No. 1, pp. 127–160.
25. Dodd, R. (2009), "Exotic derivatives losses in emerging markets: Questions of suitability, concerns for stability", International Monetary Fund, working paper series no. WP/09, Washington, DC, viewed 28 October 2012, from http://www.financialpolicy.org/ kiko.pdf
26. Dodd, R. (2012), "Markets: Exchange or over-the-counter", in International Monetary Fund: Finance and development, viewed 26 February 2013, from http://www.imf.org/external/pubs/ft/fandd/basics/markets.htm
27. Dodd, R. and Griffith-Jones, S. (2007), "Brazil's derivatives markets: Hedging, central bank intervention and regulation", in Economic Commission for Latin American and the Caribbean, viewed 20 January 2013, from http://repositorio.cepal.org/bitstream/handle/11362/5323/S0701097.pdf?sequence=1
28. Durbin, E. and Ng, D. (2005), "The sovereign ceiling and emerging market corporate bond spreads", Journal of International Money and Finance, Vol. 24 No. 4, pp. 631–649.
29. Erasmus, C. and Makina, D. (2014), "An empirical study of bank efficiency in South Africa using the standard and alternative approaches to data envelope analysis (DEA)", Journal of Economics and Behavioural Studies, Vol. 6, No. 4, pp. 310–317.
30. Financial Services Board [FSB], (2007), Circular PF No. 130 – Good governance of retirement funds, Author, Pretoria.
31. Gershkoff, A.R. (2008), "Level of measurement", in P.J. Lavrakas (Ed.), Encyclopaedia of Survey Research Methods, pp. 422–424, Sage, Thousand Oaks, CA.
32. Gill, J. and Johnson, P. (2010), Research methods for managers, 4th edn., Sage, London, UK.
33. Goodspeed, I. (2013), "Twin peaks", Financial Markets Journal, viewed 30 July 2014, from http://www.financialmarketsjournal.co.za/17thedition/printedarticles/twinpeaks.htm
34. Goyenko, R., Subrahmanyam, A. and Ukhov, A. (2011), "The term structure of bond market liquidity and its implications for expected bond returns", Journal of Financial and Quantitative Analysis, Vol. 46 No. 1, pp. 111–139.
35. Graf, S., Haertel, L., Kling, A. and Ruß, J. (2014), "The impact of inflation risk on financial planning and risk-return profiles", ASTIN Bulletin, Vol. 44 No. 2, pp. 335–365.
36. Greenwood, R. and Vayanos, D. (2010), "Price pressure in the government bond market", American Economic Review, Vol. 100 No. 2, pp. 585–590.
37. Hakansson, N.H. (1999), "The role of a corporate bond market in an economy – and in avoiding crises", China Accounting and Finance Review, Vol. 1 No. 1, pp. 1–10.
38. Hearn, B. and Piesse, J. (2012), "A reassessment of stock market integration in SADC: The determinants of liquidity and price discovery in Namibia", Applied Financial Economics, Vol. 23 No. 2, pp. 123–138.
39. Hull, J.C. (2008), Fundamentals of futures and options markets, 6th edn., Pearson, New York, NY.
40. Hull, J.C. and White, A. (2014), "Valuing derivatives: Funding value adjustments and fair value", Financial Analysts Journal, Vol. 70 No. 3, pp. 46–56.
41. International Organization of Securities Commissions [IOSCO], (2010), "OTC markets and derivatives trading in emerging markets", in Better Regulation, viewed 27 February 2013, from http://hb.betterregulation.com/external/OTC%20Markets%20and%20Derivatives%20Trading%20in%20Emerging%20Markets%20-%20Final%20Report.pdf

42. Ishak, N.M. and Bakar, A.Y.A. (2014), "Developing sampling frame for case study: Challenges and conditions", World Journal of Education, Vol. 4 No. 3, pp. 29–35.
43. Jamshidian, F. (2010), "Forward and swap measures", Encyclopaedia of Quantitative Finance, Wiley, Enschede, Netherlands.
44. Johannesburg Stock Exchange [JSE], (2011), "Bond option market forum: Minutes", in Johannesburg Stock Exchange, 22 November, viewed 2 April 2012, from http://www.jse.co.za/Libraries/ Interest_Rate_ Market_-_Volatility_Skew/04_Minutes_from_Bond_ Options_Market_Forum_-_22_November_2011.sflb. ashx
45. Johnson, L. (2014), "Geographies of securitized catastrophe risk and the implications of climate change", Economic Geography, Vol. 90 No. 2, pp. 155–185.
46. Kalač, B., Bećirović, S. and Plojović, Š. (2013), "The role of financial intermediaries in equity markets", Regional Scientific Conference with International Participation: Effects of Global Risk in Transition Countries, pp. 273–278.
47. Lawton, D. (2012), "Liquidity and the regulation of markets", paper presented at the TradeTech Liquidity conference, London, 26 April.
48. Maniar, H.M. (2007), "Impact of derivatives trading on the underlying securities: A case study on National Stock Exchange (NSE) of India", paper presented at the World Congress on Engineering, London, 2–4 July.
49. McDaniel, C., Lamb, C.W. and Hair, J.F. (2008), Introduction to marketing, Thomson South Western Publishers, Mason, OH.
50. Milne, F. (2009), "The complexities of financial risk management and systematic risks", Bank of Canada, Summer, pp. 15–29.
51. National Treasury (n.d.), Debt management report 2011/12, viewed 28 September 2014, from http://www.treasury.gov.za/publications/other/
52. National Treasury, (2012), Financial Markets Bill – Annexure C: An examination of the South African OTC derivatives markets to recommend measures for strengthening their regulatory oversight, Author, Pretoria.
53. Pacific Investment Management Company [PIMCO], (2011), "The Australian bond market: A profile for defensive investors", in PIMCO Australia Pty Ltd, viewed 12 April 2013, from http://www.rethinkyourdefence.com.au/resources/pdf/ Australian_Bond_Market_FINAL.pdf
54. Profile Group (n.d.), JSE – Interest rate market members, viewed 28 October 2013, from http://www.fmd.co.za/Index/AllByCatList.htm#SEC3 4
55. Reid, M. (2009), "Isolating a measure of inflation expectations for the South African financial market using forward interest rates", South African Journal of Economics, Vol. 77 No. 3, pp. 399–413.
56. Remenyi, D., Williams, B., Money, A. and Swartz, E. (1998), Doing research in business and management: An introduction to process and method, Sage, London, UK.
57. Sherman, M. (2009), "A short history of financial deregulation in the United States", in Center for Economic and Policy Research, viewed 19 October 2012, from http://www.cepr.net/index.php/ publications/reports/a-short-history-of-financial-deregulation-in-the-united-states
58. Stulz, R.M. (2005), "Demystifying financial derivatives", The Milken Institute Review, Vol. Third quarter, pp. 20–31.
59. Van Heerden, C. and Van der Westhuizen, G. (2008), "The cost of banking services and the technical efficiency of a South African bank", The Journal for Transdisciplinary Research in Southern Africa, Vol. 4 No. 2, pp. 401–417.
60. Wuite, R. (2009), The South African dictionary of finance, Frontrunner, South Africa.

THE IMPACT OF FISCAL AND MONETARY POLICIES ON UNEMPLOYMENT PROBLEM IN NIGERIA (MANAGERIAL ECONOMIC PERSPECTIVE)

Nick Attamah*, Igwe Anthony**, Wilfred I. Ukpere***

Abstract

This paper investigates the impact of fiscal and Monetary Policies on Unemployment Problem in Nigeria and covers the periods 1980 to 2013. To achieve this, fiscal policy was captured here by government expenditures and revenues respectively while monetary policy was proxied by broad Money Supply (M2), Interest and Exchange rates respectively. The methodology adopted was econometric analysis employing OLS techniques and unit roots of the series were examined using the Augmented Dickey-Fuller after which the co-integration tests was conducted using the Engle Granger approach. Error correction models were estimated to take care of the short run dynamics. It was found that while government expenditure had a positive relationship with unemployment problem in Nigeria, the result of government revenue was negative and insignificant on unemployment problem. For monetary policy, it was found that money supply and exchange rate had positive and significant impact while interest rate has only a positive relationship on unemployment problem in Nigeria. This meets the a priori expectation. The study also revealed that increases in interest and exchange rates escalate unemployment by increasing cost of production which discourages the private sector from employing large workforce. On the other hand, national productivity measured by real GDP had a negative and significant impact on unemployment rate in Nigeria. This paper recommends that for an effective combat to unemployment problem in Nigeria, there should be a systematic diversion of strategies, thus more emphasis should be laid on aggressively pursuing entrepreneurial development and increased productivity. Again government should aggressively focus on investment, employment generation and economic growth that has mechanism to trickle does to the masses.

Keywords: Fiscal, Monetary, Unemployment Rate, Unit Root, Productivity Global Recession

**Economics Department, Enugu State University of Science and Technology (ESUT)*
***University of Nigeria, Enugu Campus*
****Department of Industrial Psychology and people Management, Faculty of Management, University of Johannesburg, South Africa*

1 Introduction

The economic thinking before 1930, generally referred to as the classical economics propounded that the economy will always be at full employment state without inflation. This is so because for them, the demand for labour will always equal the supply of labour at the prevailing money wage rate. For the classical economists, if for any reason, there was an in increase in labour supply the money wage will fall and more workers would be employed. Similarly, if there is a shortage of workers the money wage will rise thereby eliminating the shortage. However, with the great depression of the 1930s, this classical theory could not hold as there was wide spread unemployment.

Today, unemployment may be viewed as one of the most intractable problems facing Nigeria since 1960 and climaxing in these millennium years. It has become a cankerworm that is now eating deep into the fabric of the Nigerian economy. The existence of high unemployment in any economy is a source of concern to policy makers as well as the general citizenry. According to Layard, Nickell & Jackman (1994) unemployment generally reduces output and aggregate income. It increases inequality since the unemployed lose more than the employed. It erodes human capital and involves psychic costs. Though unemployment increases leisure, the pain of rejection largely offsets the value of this. Those who are unemployed sometimes feel as if the society does not need them.

According to Englama (2001), the issue of persistent unemployment is now frightening in Nigeria considering the fact that it is widening poverty, misery, and social unrest, ethnic cum religious crisis, robbery, kidnappings, terrorism and other social vices. These have posed a great challenge to policy makers/planners, human resource experts and persons dealing with unemployment programmes, planning and implementations. In recognition of the crucial role of and the need for manpower

development, the Federal Government of Nigeria appointed the Ashby commission in 1959 to look into Nigeria's needs in the field of post school certificate and higher education during the two decades 1960-1980. It is striking to note that the expiration of the planner's period covered by the report marked the beginning of mass unemployment in Nigeria.

Since the oil boom of the early 1970's the revenue base of Nigeria has depended largely on the oil sector which has provided more than 96 percent of total export earnings. The oil boom provided the opportunity for government to initiate gigantic expenditure programmes which reduced the rate of unemployment. Shortly after that, starting from mid-1981, the World oil market began to collapse and with it, a traumatic economic crisis emerged in Nigeria. The government then had to borrow both from internal and external sources which resulted into fiscal deficit. The fiscal deficit created economic instability with high inflation rate which reduced the gains previously made in reducing unemployment. In addition, inappropriate and ineffective policies of the past such as the Economic Stabilization (1982) and Economic Emergency (1985) measures aggravated the economic quagmire. In effect, these austerity measures dramatically reduced supply of new materials and spare parts to the import dependent industrial sector, resulting in extensive plant closure, substantial drop in capacity utilization and retrenchment of many workers (Anyanwu, Oyefusi , Oaikhenah, Dimowo, 1997).

In an effort to encourage employment generation using monetary policy, interest rates were liberalized (deregulated) and were also controlled on several occasions. The failure of these polices and the onward escalation of unemployment problem necessitated the government to introduced Structural Adjustment Programme (SAP) in July 1986 and also National Directorate of Employment which was mainly to encourage self-employment by granting loans to prospective individuals who want to be self-employed. The government also introduces the 6-3-3-4 system of education to arrest the problem of unemployment in Nigeria. Despite the introduction of these novel programmes, unemployment problem has remained intractable probably due to increases in population and the Proliferation of Secondary and Tertiary education in Nigeria. This paper investigates the impact of fiscal and monetary policies on unemployment problem in Nigeria from 1980 to 2013, which is a period of thirty three years which gives enough degree of freedom for a reliable estimation result.

2 Aim of the paper

The paper aims to investigate the impact of fiscal and monetary policies on unemployment problem in Nigeria from 1980-2013.

3 Structure of the paper

The paper is organized as follows: Section 1 is the introduction, while section 2 is the research objectives. Section 3 is the structure of the paper, while in section 4, we review pertinent literatures. Section 5 discusses the research methodology and data sources, while Section 6 is the data analysis and discussion of results. Conclusion and recommendations are contained in section 7.

4 Literature review

Adeyemi (2000) undertakes an analysis of the impact of Development plans on employment generation and offers some policy lessons. He opines that all post-independence plans as always had employment generation as one of their cardinal objectives. In addition, efforts have always been made through the national Manpower Board to ensure that employment sensitive manpower programmes are addressed. A number of key manpower development institutions were established between 1971 and 1986 to address some areas of manpower lapses. The challenge is that their programmes should be upgraded to meet the managerial and skill requirements to cope with challenges of globalizing world and it attendant competitive pursues. He concluded by arguing that these organization can play significant role to enhance knowledge utilization in the economy.

Damachi (2001) in his study of past policy measures for solving unemployment problems in Nigeria suggests that there is a strong need for institutional collaboration and improved coordination of policy measures for dealing with unemployment. He stated that while there are some discernable lapses, the overall policy direction for employment appears to be adequate. According to him, what is required is the political will to pursue the policy measures backed by adequate steps to make the polices work as well as transparency in programme implementation.

Okekukola (2006) in his study recommends that given the level of unemployment in Nigeria, the development of entrepreneurial skills and initiatives should be of paramount important especially in the higher education sector. This will facilitate employability of graduates who will increasingly be called upon to be not only job seekers, but above all to become job creators. He opined that emphasis should be placed on facilitating the acquisition of skills, competencies and ability which are required by employees of labour. He concluded that government has a pivotal role to play in an effort at finding real and lasting solution to this malaise.

Kahn (1993) offers some explanation for the high rate of unemployment in the United States of America (USA). Technological advancement is one explanation. The computers, which were introduced into the production process, were effectively utilized to their full capacity by 1990. The complete

absorption of computer technology into the factories may have resulted in a drastic cut in labour force. Again, the global recession has contributed somewhat to the unemployment rate now ravaging the US economy. The recession showed the economies of US trading partners and consequently reduced demand for its experts.

Yesufu (1984) in his study argues that higher labour productivity can reduce unemployment. This can be explained by the marginal productivity principle which says that employers will hire more labour up to the point where the value of marginal product of labour equals marginal labour cost (Okojie, 1995). Thus, an increase in marginal productivity labour indicates that employment of labour would have to increase in compliance with the marginal productivity principle. However, Diacharbe (1991) hold a contrary view which says higher productivity of labour may increase unemployment. This is because fewer workers can be used to achieve a given level of output, while the redundant workers may be laid off. According to Prokopenko (1992) a decline in the productivity of labour would always lead to economic deterioration and a consequent rise in unemployment rate.

Sanusi (1997) in his study titled stimulating investment through interest rate management reported that interest rate has positive relationship with unemployment that is, a lower interest rate encourages private investment spending which will increase the demand for labour and reduce unemployment. According to him, high interest rate (Prime Lending Rate) has characterized the Nigerian economy over the years and this has adversely affected the manufacturing sector which ought to significantly reduce unemployment. He concluded by urging the authorities to reduce the prime lending rate as this could reduce unemployment problem in the economy.

Adebusuyi (1997) studied the performance evaluation of small and medium enterprises in Nigeria. He recommends that SMEs will provide an engine for growth and prosperity and thereby create job opportunities. He cited example of Mauritius where the potential of SMEs for job creation was well demonstrated in the early 1980s, when economic recession led to high unemployment, however, SMEs reduced unemployment from 21.0 percent in 1983 to 1.6 percent in 1996.

Also in Thailand, SMEs constitute more than 90 percent of the total number of establishments in the manufacturing sector. They employ about 65 percent of the industrial workers and constitute about 47 percent of the total manufacturing value added. The Thai government uses SMEs as instrument to create employment, to harness and effectively use given natural resources and to narrow income gap. SMEs played critical role in Malaysia's industrialization programme through the strengthening of both forward and backward industrial linkages. It is the same story in Pakistan where SMEs constitute 90 percent of business and accounts for 80 percent of total employment and 30 percent of Gross Domestic Product (GDP).

Borishade (2001) in his study of restructuring the educational system as a long term solution to the unemployment problem in Nigeria concluded that education is the key with which to unlock the economic potential of the people as it empowers the individual to improve himself as well as equip him to participate in, contribute to, and drive benefits from the national economy. To him, the reinvigoration of the technical and vocational education is a worthy step in this direction. He however, concluded that the pluralistic nature of Nigerian society makes it imperative that all sectors of the country have to be carried along in the developmental process.

5 Methodology and data sources

5.1 Theoretical framework

From the reviewed literature, we illustrated that unemployment problem depends on a variety of factors. This study anchors on five identified factors to explain the unemployment problems in Nigeria. These factors include:

a. Government revenue,
b. Government expenditure,
c. Interest rate,
d. Money supply M_2, and
e. Exchange rate

Changes in government revenue and government expenditure are indicators of fiscal policy, while the monetary policy is proxied by changes in interest rate, money supply and exchange rate. An increase in government expenditure, all things being equal, leads to expansion in the production and workforce (Kahn, 1993, Keynes 1936). For studies that involve measurement of variables such as in this study, analytical method is the most appropriate method to be used. Hence this study is analytical in nature and econometric analysis was used employing OLS technique. Secondary data that captures unemployment, fiscal and monetary policy variables in Nigeria for the period 1980-2013 were also used in the study and were extracted from a secondary source (Central Bank of Nigeria Statistic Bulletin). The scope was limited to 2013 because it is the most recent annual data available based on the variables used in the analysis for the time of the research.

5.2 Model specification

The model adopted for the paper assumes an underlying relationship between unemployment, fiscal and monetary policies. The belief was informed by the Keynesian proposition that unemployment can be controlled and combated by the use of fiscal and monetary policy tools.

The model is specified implicitly below:

$$UN = f(GEX, GR, MS, IR, EXR, GDP)$$

The econometric form of equation (1) above is specified thus:

$$UNr = \Omega_0 + \Omega_1 GEX + \Omega_2 GR + \Omega_3 MS + \Omega_4 IR + \Omega_5 EXR + \Omega_6 GDP + \varepsilon_t$$

Where:
UN = Unemployment Rate
GEX = Government Expenditures
GR = Government Revenues
MS = Money Supply
IR = Interest Rate
EXR = Exchange Rate
GDP = Productivity Variable
ε_t = Stochastic Error Term

5.3 Unit root test

The estimation of variable-series that are non-stationary will thus lead to estimates that are spurious and thus render the coefficients unreliable for policy prescription and usage. This entails that the investigation will thus carry out the conventional unit-root tests on each of the variables to be used in this analysis. The stationarity test will be carried out with the application of Augmented-Dickey Fuller Statistic.

The following three models represent pure random walk, random walk with drift and random walk with drift and trend used in Augmented Dickey Fuller tests:

$$\Delta\psi_t = \Omega\psi_{t-1} + \sum_{i=1}^{p} \beta_i \Delta\psi_{t-1} + \varepsilon_t$$

$$\Delta\psi_t = \alpha_0 + \Omega\psi_{t-1} + \sum_{i=1}^{p} \beta_i \Delta\psi_{t-i} + \varepsilon_t$$

$$\Delta\psi_t = \alpha_0 + \Omega\Psi + \beta_2 t + \sum_{i=1}^{p} \beta_i \Delta\psi_{t-1} + \varepsilon_t$$

where: $\Omega = (\lambda - 1)$ The null hypothesis is $H_0 : \Omega = 0$ and the alternative hypothesis is $H_a : \Omega < 0$. If ADF test statistic (t-statistic of lagged dependent variable) is less than the critical value, we reject the null hypothesis and conclude that the series is stationary (there is no unit root).

5.4 Co-integration test

The co-integration test will be carried out which allows for the estimation of a long-run equilibrium relationship. Simply put, one can argue that various non-stationarity time series are co-integrated when their linear combination are stationary. Stationary derivations from the long run are allowed in the short run. Economically speaking two variables can only be co-integrated if they have long-term or equilibrium relationship between them.

5.5 Error correlation mechanism (ECM)

The error correlation mechanism is employed to tie the short-run dynamic behaviours of a variable to its long-run value. The error correlation mechanism (ECM), which was first used by Sargan and later popularized by Engle and Granger (1987) correct for disequilibrium. Given these dynamics, Engle and Granger suggested that adjustment should be involved through the (iterative) process to obtain a more parsimonious model. The ECM is stated as:

$$\Delta Y_t = \theta_0 + \theta_1 z_{t-1} + \sum \theta_{2i} \Delta X_{t-1} + \sum \theta_{3i} \Delta Y_{t-1} + \varepsilon_t$$

Where Δ denotes the first order time difference (i.e $\Delta y_t = y_t - y_{t-1}$) and where ε_t is a sequence of independent and identically distributed random variables with mean Zero and variance. Furthermore, they prove the converse result that an ECM generates cointegrated series.

6 Discussion of results

6.1 Unit root result

For a guide to an appropriate specification of the regression equation, the characteristics of the time series data used for estimation of the model were examined to avoid spurious regression. We begin by determining the under lying properties of the process that generate our time series variables, that is whether the variables in our model were stationary or non-stationary.

Macroeconomic data often possess stochastic trends that can be removed by differencing the variables. We therefore employ the Augmented Dickey Fuller (ADF) to test the order of integration of the variables.

The ADF results are displayed on the table below:

Table 1. Unit root result

Variables	ADF-statistics	Critical value	Order of inegration
Unemployment	-3.954425	-1.9521	I(1)
GEX	-2.460021	-1.9517	I(0)
GR	-2.579226	-1.9526	I(2)
MS	-2.753835	-1.9517	I(0)
IR	-5.462748	-2.9591	I(1)
EXR	-3.181917	-1.9521	I(1)
GDP	-8.538947	-1.9526	I(1)
RESIDUALS	-2.761750	-1.9517	I(0)

Note: the critical value is based on 5% level of significance

The stationary test result above shows that unemployment rate, Interest Rate (IR), Exchange Rate (EXR), and measure of Productivity (GDP) are stationary at first difference I(1), Government Revenue (GR) is stationary at second difference I(2) and Money Supply (MS) and residuals are stationary at level I(0).

Table 2. Regression results

Dependent Variable: D(UNR)
Method: Least Squares
Date: 01/28/15 Time: 17:30
Sample(adjusted): 1981 2013
Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.947953	3.270256	-0.901444	0.3760
LOG(GOVTEXP)	0.349326	0.258793	1.349829	0.1892
DLOG(GOVTREV)	1.427962	1.130467	1.263161	0.2182
DLOG(MS)	1.235099	0.826547	1.494288	0.1476
DLOG(IR)	3.936966	2.181042	1.805085	0.0831
DLOG(EXR)	1.718592	1.940550	0.885621	0.3843
DLOG(GDP)	-9.032251	3.449686	-2.618282	0.0148
ECM(-1)	-0.455944	0.168034	-2.713402	0.0119
R-squared	0.455916	Mean dependent var		0.603030
Adjusted R-squared	0.303572	S.D. dependent var		3.259456
S.E. of regression	2.720090	Akaike info criterion		5.046424
Sum squared resid	184.9722	Schwarz criterion		5.409213
Log likelihood	-75.26599	F-statistic		2.992678
Durbin-Watson stat	2.489825	Prob(F-statistic)		0.020021

The regression result above shows that both the fiscal and monetary policy variables have no significant impact on unemployment problem in Nigeria. As clearly shown, their corresponding t-statistics yielded values absolutely less than 2 and with probability values exceeding 0.05 (5%). However the productivity variable which is measured with GDP is found to be statistically significant yielding a t-statistics of -2.713402 and a coefficient value of -9.032251. The negative coefficient however shows an inverse relationship that exists between national productivity and unemployment. This clearly conforms to economic growth a priori expectation because an increase in national productivity is expected to reduce the level of unemployment in the economy and vice-versa. The R-Squared Statistic which yielded 0.455916 entails that the explanatory power of the independent variables [Fiscal and Monetary Policy variables] is below average and hence not considered high and significant. The F-statistic which yielded 2.992678 and is seen to be less than absolute 3 entails that the test is statistically insignificant at the entire regression plane.

6.2 Cointegration test result/ECM

The Engle-Granger cointegration test results confirm the existence of long-run relationship among the variables by the stationarity of the residuals at level form, a seen in the appendix.

The results of the ECM shows that the short-run dynamics restores back to long-run equilibrium at 45.5%. This shows that the speed of the adjustment to long-run equilibrium is not fast but slightly below average.

7 Conclusion

The paper investigated the impact of fiscal and monetary policies on unemployment problem in Nigeria ranging from 1980-2013. To achieve this, we captured Fiscal Policy with Government Expenditures and Government Revenue, and the monetary policy was proxied with Money Supply (M_2), Interest Rate and exchange rate. The estimation of the model was estimated with the application of linear regression. It was found that while government expenditure had a positive relationship with unemployment problem in Nigeria, the result of government revenue was negative and insignificant on unemployment problem. For monetary policy, it was found that money supply and exchange rate had positive and significant impact while interest rate has only a positive relationship on unemployment problem in Nigeria. On the other hand, national productivity measured by real GDP had a negative and significant impact on unemployment rate in Nigeria.

In conclusion, fiscal and monetary policies have been more effective in general economic growth as evidenced in steady growth of the GDP. This is however ineffective in some other areas like unemployment generation, exchange rate etc. This paper therefore recommends that for an effective combat of unemployment problem in Nigeria, there should be a systematic diversion of strategies, thus more emphasis should be laid on aggressively pursuing entrepreneurial development and increased productivity. Again government should aggressively focus on investment, employment generation and economic growth that has mechanism to trickle down to the masses. More than that, foreign and domestic investors should be encouraged to invest in the key sectors like Agriculture and Manufacturing to help in diversifying the economy and hence increase the employment generation.

References

1. Anyanwu J.C, Oyefusi A, Oaikhenah. H, Dimowo F.A (1997), The structure of the Nigerian Economy. Onitsha, Joanee Educational Publishers Ltd.
2. Adeyemi J.O (1996) Impact of Development Plans on Employment Generation; Some Policy Lessons; in J.U Umo ed. (1996), Towards Full Employment Strategy in Nigeria: Lagos, Nigeria.
3. Borishade A, B (2001), Restructuring the Educational System as a Long-term Solution to the Unemployment Problem in Nigeria, in CBN Bullion Vol. 25 No 4.
4. Damachi N.A (2001) Evaluation of Past Policy Measures for Solving Unemployment Problems, in CBN Bullion vol. 25 No 4.
5. Diacharbe O (1991), "Productivity Data and Economic Growth in Nigeria, in Productivity for Self-Reliance and Excellence", published by National Productivity Center (NPC), pp. 65-69.
6. Englama A (2001), Unemployment Concepts and Issues in CBN Bullion vol.25 No. 4
7. Engle R.F and Granger, C.W (1987), "Co-integration and Error – Correction: Representation, Estimation and Testing", Econometric vol 55 No. 2. Pp.251-276.
8. Kahn G.A (1993), "Sluggish Job Growth: Is Rising Productivity or an Anemic Recovery of Blame?", Economic Review, Federal Reserve Bank of Kansas City, Vol. 78 No.3 pp. 7-25
9. Keynes, J.M (1993) General Theory of Employment, Interest and Money, New York, Harcourt, Brace and World, Inc.
10. Layard R, Nickell S and Jackman R (1994), The Unemployment Crisis, Oxford, Oxford University Press.
11. NDE (2000), Promoting Support for Graduate Employment Creation PRS 2000
12. Okekukola P, (2006), Access Retention and Attrition as Key Issues in attainment of Education for all in Nigeria – Paper presented at a National Conference on "Education for All by the Year 2000".
13. Prokopenko, J (1992) Productivity Management, A Practical Handbook H.O.Geneva
14. Sanusi J.O (2001) The Nigerian Economy: Growth, Productivity and the Role of Monetary Policy in CBN Bullion vol. 25 No. 4
15. Sanusi J.O (1997), "Stimulating Investment Through Interest Rate Management" The Economy, Nov. 10 p.18
16. Yesufu T.M (1984), "The Dynamics of Industrial Relations, The Nigerian Experience". (Ibadan, University Press).

Appendix A

Table A.1. Unit root test results

ADF Test Statistic	-3.954425	1% Critical Value*	-2.6395
		5% Critical Value	-1.9521
		10% Critical Value	-1.6214

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(UNR,2)
Method: Least Squares
Date: 01/28/15 Time: 15:50
Sample(adjusted): 1983 2013
Included observations: 31 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(UNR(-1))	-1.150243	0.290875	-3.954425	0.0005
D(UNR(-1),2)	-0.112017	0.181616	-0.616776	0.5422

R-squared	0.657798	Mean dependent var	-0.083871
Adjusted R-squared	0.645998	S.D. dependent var	5.473822
S.E. of regression	3.256821	Akaike info criterion	5.261721
Sum squared resid	307.5995	Schwarz criterion	5.354236
Log likelihood	-79.55667	Durbin-Watson stat	2.009832

ADF Test Statistic	-2.460021	1% Critical Value*	-2.6369
		5% Critical Value	-1.9517
		10% Critical Value	-1.6213

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GOVTEXP)
Method: Least Squares
Date: 01/28/15 Time: 15:52
Sample(adjusted): 1982 2013
Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GOVTEXP(-1)	-0.232074	0.094338	-2.460021	0.0199
D(GOVTEXP(-1))	1.430340	0.630954	2.266950	0.0308

R-squared	0.181586	Mean dependent var	13119.38
Adjusted R-squared	0.154305	S.D. dependent var	727203.5
S.E. of regression	668748.4	Akaike info criterion	29.72466
Sum squared resid	1.34E+13	Schwarz criterion	29.81627
Log likelihood	-473.5946	Durbin-Watson stat	1.980316

ADF Test Statistic	-2.579226	1% Critical Value*	-2.6423
		5% Critical Value	-1.9526
		10% Critical Value	-1.6216

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GOVTREV,3)
Method: Least Squares
Date: 01/28/15 Time: 15:54
Sample(adjusted): 1984 2013
Included observations: 30 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GOVTREV(-1),2)	-12.03372	4.665633	-2.579226	0.0154
D(GOVTREV(-1),3)	5.485619	2.549482	2.151660	0.0402

R-squared	0.172587	Mean dependent var	3004682.
Adjusted R-squared	0.143036	S.D. dependent var	16873433
S.E. of regression	15620130	Akaike info criterion	36.03036
Sum squared resid	6.83E+15	Schwarz criterion	36.12377
Log likelihood	-538.4554	Durbin-Watson stat	0.648379

Table A.1. Unit root test results (continued)

ADF Test Statistic	-2.753835	1% Critical Value*	-2.6369
		5% Critical Value	-1.9517
		10% Critical Value	-1.6213

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MS)
Method: Least Squares
Date: 01/28/15 Time: 15:56
Sample(adjusted): 1982 2013
Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MS(-1)	-0.621710	0.225762	-2.753835	0.0099
D(MS(-1))	-0.191698	0.180941	-1.059450	0.2979

R-squared	0.404675	Mean dependent var	206627.0
Adjusted R-squared	0.384831	S.D. dependent var	12773747
S.E. of regression	10018799	Akaike info criterion	35.13829
Sum squared resid	3.01E+15	Schwarz criterion	35.22989
Log likelihood	-560.2126	Durbin-Watson stat	2.021613

ADF Test Statistic	-5.462748	1% Critical Value*	-3.6576
		5% Critical Value	-2.9591
		10% Critical Value	-2.6181

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(IR,2)
Method: Least Squares
Date: 01/28/15 Time: 15:57
Sample(adjusted): 1983 2013
Included observations: 31 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(IR(-1))	-1.736216	0.317828	-5.462748	0.0000
D(IR(-1),2)	0.188093	0.185979	1.011370	0.3205
C	0.647724	0.739960	0.875350	0.3888

R-squared	0.739312	Mean dependent var	0.016129
Adjusted R-squared	0.720691	S.D. dependent var	7.698120
S.E. of regression	4.068432	Akaike info criterion	5.736158
Sum squared resid	463.4598	Schwarz criterion	5.874931
Log likelihood	-85.91045	F-statistic	39.70400
Durbin-Watson stat	1.902380	Prob(F-statistic)	0.000000

ADF Test Statistic	-3.181917	1% Critical Value*	-2.6395
		5% Critical Value	-1.9521
		10% Critical Value	-1.6214

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EXR,2)
Method: Least Squares
Date: 01/28/15 Time: 16:12
Sample(adjusted): 1983 2013
Included observations: 31 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(EXR(-1))	-0.863764	0.271460	-3.181917	0.0035
D(EXR(-1),2)	-0.099807	0.195899	-0.509482	0.6143

R-squared	0.454775	Mean dependent var	0.841184
Adjusted R-squared	0.435974	S.D. dependent var	19.42598
S.E. of regression	14.58923	Akaike info criterion	8.260786
Sum squared resid	6172.527	Schwarz criterion	8.353301
Log likelihood	-126.0422	Durbin-Watson stat	1.947179

Table A.1. Unit root test results (continued)

ADF Test Statistic	-8.538947	1% Critical Value*	-2.6423
		5% Critical Value	-1.9526
		10% Critical Value	-1.6216

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GDP,3)
Method: Least Squares
Date: 01/28/15 Time: 16:13
Sample(adjusted): 1984 2013
Included observations: 30 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GDP(-1),2)	-2.772168	0.324650	-8.538947	0.0000
D(GDP(-1),3)	0.668281	0.178886	3.735786	0.0008
R-squared	0.854573	Mean dependent var		183649.2
Adjusted R-squared	0.849379	S.D. dependent var		4064068.
S.E. of regression	1577261.	Akaike info criterion		31.44462
Sum squared resid	6.97E+13	Schwarz criterion		31.53803
Log likelihood	-469.6693	Durbin-Watson stat		1.998242

Table A.2. Residual test of co-integration

ADF Test Statistic	-2.761750	1% Critical Value*	-2.6369
		5% Critical Value	-1.9517
		10% Critical Value	-1.6213

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RESID01)
Method: Least Squares
Date: 01/28/15 Time: 17:16
Sample(adjusted): 1982 2013
Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	-0.578919	0.209620	-2.761750	0.0097
D(RESID01(-1))	-0.116542	0.184021	-0.633311	0.5313
R-squared	0.336241	Mean dependent var		0.114842
Adjusted R-squared	0.314115	S.D. dependent var		3.914481
S.E. of regression	3.241901	Akaike info criterion		5.250658
Sum squared resid	315.2976	Schwarz criterion		5.342267
Log likelihood	-82.01053	Durbin-Watson stat		1.966954

Table A.3. Regression result using ECM

Dependent Variable: D(UNR)
Method: Least Squares
Date: 01/28/15 Time: 17:30
Sample(adjusted): 1981 2013
Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.947953	3.270256	-0.901444	0.3760
LOG(GOVTEXP)	0.349326	0.258793	1.349829	0.1892
DLOG(GOVTREV)	1.427962	1.130467	1.263161	0.2182
DLOG(MS)	1.235099	0.826547	1.494288	0.1476
DLOG(IR)	3.936966	2.181042	1.805085	0.0831
DLOG(EXR)	1.718592	1.940550	0.885621	0.3843
DLOG(GDP)	-9.032251	3.449686	-2.618282	0.0148
ECM(-1)	-0.455944	0.168034	-2.713402	0.0119
R-squared	0.455916	Mean dependent var		0.603030
Adjusted R-squared	0.303572	S.D. dependent var		3.259456
S.E. of regression	2.720090	Akaike info criterion		5.046424
Sum squared resid	184.9722	Schwarz criterion		5.409213
Log likelihood	-75.26599	F-statistic		2.992678
Durbin-Watson stat	2.489825	Prob(F-statistic)		0.020021

EXAMINING THE GENDER INFLUENCE ON EMPLOYEES' WORK ENGAGEMENT WITHIN A SOUTH AFRICAN UNIVERSITY

MC Tshilongamulenzhe*, N Takawira**

Abstract

Organisations should prioritize and promote employees' work engagement if they are to effectively realise their goals and objectives. This study seeks to examine the influence of gender on employees' levels of work engagement within a South African university. A quantitative non-experimental cross-sectional survey design was used to collect data from 154 university employees who were purposefully selected to participate in this study. Data were analysed using SPSS software (version 23.0) and the findings show no statistically significant gender differences amongst employees with regard to their levels of work engagement. The findings provide scope for further research in South Africa which should examine ethnic and cultural connotations related to gender and further investigate how these influence employees' work engagement.

Keywords: Gender Influence, Work Engagement, South Africa

**Corresponding author. Department of Human Resource Management, University of South Africa, P.O. Box 392, UNISA, 0003, South Africa*
Tel: +2712 429 3724
Fax: +2786 642 2062
***Department of Human Resource Management, University of South Africa, P.O. Box 392, UNISA, 0003, South Africa*
Tel: +2712 429 4045
Fax: +2786 690 5927

1 Introduction

Competitive organisations depend on skilled, competent and engaged employees in order to develop, maintain and sustain various strategies which enable effective production and delivery of goods and services. Such professionals are in short supply globally and South Africa is not spared from the prevailing talent war. Sharp (2011) indicated that lack of skills will make it difficult for South Africa to productively tap into viable economic opportunities. However, as work is becoming more complex, employers are looking for more specific skills, thus leading to an increasing mismatch between demand and supply (Zieminski, 2009). There are diverse scholarly opinions suggesting that the global talent crunch will not be alleviated anytime soon, but would almost certainly continue for some years to come (Keeffe, 2010; Neumark, Johnson and Mejia, 2011).

As the global talent crunch deepens, many organisations will find it difficult to hold onto their available talent. Some of these organisations are already awakening to the need to grow and develop the intellectual capital of their employees in order to compete successfully in an increasingly demanding global economy (Burke and El-Kot, 2010; Lawler, 2008). However, developing employees without strategies to retain them and keep them engaged seems to be a futile exercise. In reality, the changing nature of work and the shifting demographics of the workforce today call upon employers to be more focused on retaining their talent and keeping such talent fully engaged. This is mainly because work engagement is claimed to have positive consequences for both the employees and the organisations (Saks, 2006; Jeung, 2011; Bhatnagar, 2012). As a result, organisations may not effectively realise their growth objectives without the competent and engaged employees to make it happen (Seegers, 2012). Oberholster (2014) suggest that engagement happens when employees choose to invest the best of who they are in pursuit of a worthy mission. Employees choose to invest when they perceive that they will get a desired return, that is, when an organisation invests back in them.

2 Theoretical perspective and trends regarding work engagement

Work engagement is regarded as one of the hottest topics in management (Welbourne, 2007) and certainly something worth promoting (Taris, Cox and Tisserand, 2008). It is a multi-dimensional construct that involves behavioural as well as attitudinal characteristics which lead to a positive experience in the workplace. Barkhuizen (2014) advocates that organisations often have three types of employees: the engaged, the not-engaged and the actively disengaged.

The role of an effective leader is to ensure an environment that can facilitate the engaged, romance the not-engaged and make the actively disengaged fall head over heels so that they can add value to the organisation (Barkhuizen, 2014). Helping employees understand the importance of their job in achieving organisational success is critical for engagement. Managers must always remember that employee work engagement and organisational success are not polar opposites.

Engagement will, in all likelihood, have a positive impact on the organisation's bottom line. This is because, engaged employees become more attached to their organisations and would have a lower propensity to leave it (Schaufeli and Bakker, 2004), display more proactive behaviour (Salanova and Schaufeli, 2008), perform better (Bakker, Demerouti and Verbeke, 2004), obtain higher objective financial returns for the business (Xanthopoulou, Bakker, Damerouti and Schaufeli, 2007), and show less sickness absenteeism (Schaufeli, Bakker and Van Rhenen, 2009). Bakker and Demerouti (2007) and Bakker (2009) argue that there are four reasons that make engaged workers perform better than non-engaged workers. First, engaged workers experience positive emotions such as happiness, enthusiasm and joy. Second, engaged workers have better health (Schaufeli and Bakker, 2004; Schaufeli, Taris and Rhenen, 2008). Third, engaged workers are able to create their own job and personal resources (Xanthopoulou et al., 2007). Lastly, engaged workers often transfer their engagement to others (Bakker, Gierveld and Van Rijswijk, 2006).

Kahn (1990) describes work engagement as the employment of oneself physically, cognitively and emotionally during role performance. It is argued in the literature that work engagement cures most organisational problems (Banihani, Lewis and Syed, 2013). According to Schaufeli, Salanova, Gonzalez-Roma and Bakker (2002, p.71), work engagement entails 'a positive, fulfiling and work-related state of mind that is characterized by three components: vigour, dedication and absorption. Vigour relates to elevated levels of energy and mental resilience when a person is occupied with a work activity, the willingness to invest effort in one's work activity and showing perseverance when faced with difficulty. Thus, employees who feel great vigour at work are highly motivated by their jobs and are likely to remain persistent when encountering difficulties (Mauno, Kinnunen and Ruokolainen, 2007). Dedication is characterised by a strong involvement and pride in one's work, coupled with a sense of significance, passion and inspiration. Absorption is characterised by being so completely focused and absorbed in one's work that time passes rapidly and one has difficulty detaching oneself from a work activity. Time passes quickly and one forgets everything else that is around.

Nevertheless, there is substantial evidence in the literature which assumes that work engagement is gender-neutral and that women and men can equally demonstrate their engagement in the workplace (Banihani et al., 2013). The absence of gender reference in the organisational research does not imply gender neutrality, but instead, that a male perspective is tacitly taken (Calas and Smircich, 1992). According to Martin (2000) by ignoring gender in organisational research and theory, researchers are contributing towards the perpetuation of inequality in the workplace. This research seeks to examine the influence of gender on employees' work engagement in a South African university.

In order to understand work engagement as a construct, this study will be guided by Kahn's (1990) theoretical foundation. Several studies on work engagement drew on Kahn's theoretical foundation and his work is regarded as the foundation for work engagement literature and the most influential one (Rothbard, 2001; May, Gilson and Harter, 2004; Koyuncu, Burke and Fiksenbaum, 2006; Saks, 2006; Christian and Slaughter, 2007; Bakker, Schaufeli, Leiter and Taris, 2008; Saks, 2008; Kim, Shin and Swanger, 2009; Simpson, 2009; Rich, Lepine and Crawford, 2010; Christian, Garza and Slaughter, 2011). Kahn (1990) embraced the notion of psychological presence when he studied and introduced the construct of engagement in organisational literature. He presented work engagement as the psychological experience of work and work contexts which influence the process by which people present or absent their selves during role performance.

Kahn grounded his research work on empirical research and existing theoretical frameworks from a number of pioneering scholars (sociologists such as Merton (1957) and Goffman (1961); psychologists such as Freud (1922); and group theorists such as Bion (1961), Slater (1966) and Smith and Berg (1987)) who supported the idea that "people are inherently ambivalent about being members of ongoing groups and systems and seek to protect themselves from both isolation and engulfment by alternately pulling away from and moving towards their memberships. These pulls ad pushes are peoples' calibrations of self-in-role, enabling them to cope with both internal ambivalences and external conditions" (Kahn, 1990, p. 694). He later describes this process of calibrations of self-in-role as personal engagement and disengagement.

Personal engagement is defined as "the harnessing of organisation members' selves to their work roles" (Kahn, 1990, p. 964). People express themselves physically, cognitively and emotionally during role performance when they become engaged, and they withdraw and separate themselves physically, cognitively and emotionally from work roles when they become disengaged. Kahn's (1990) further study on conditions of work that affect people engagement found three psychological antecedents: meaningfulness, safety and availability. Psychological

meaningfulness refers to "a feeling that one is receiving a return on investment of one's self in a currency of physical, cognitive or emotional energy (Kahn, 1990, p. 704). Aspects which influence psychological meaningfulness include task characteristics, role characteristics and work interactions. Psychological safety is "feeling able to show and employ one's self without fear of negative consequences to self-image, status or career (Kahn, 1990, p. 708). Aspects which influence the psychological safety dimension include interpersonal relationships, groups and intergroup dynamics, management styles and organisational norms. Psychological availability refers to a "sense of having the physical, emotional or psychological resources to personally engage at a particular moment (Kahn, 1990, p. 714). Aspects which influence psychological availability include physical and emotional energies, insecurity and external life activities.

These three psychological antecedents were tested by May et al. (2004) whose findings confirmed their positive relationship with work engagement. Kahn's (1990) perspective provides a holistic view on work engagement and considers both personal and organisational resources in predicting work engagement. It also considers work engagement as a distinct concept. Rich et al. (2010) confirmed that Kahn's perspective provides a sound conceptual base for research on engagement, hence the decision to embrace it as a theoretical foundation in this study. The foregoing discussion provides positive evidence for enhancing work engagement of employees in an organisation. The question that remains to be answered in this study is: does gender influence employees' work engagement in a target South African university?

3 The interface between gender and work engagement

The Boston College conducted a study on age and generations in relation to work engagement, and gender was found to be one of the overall drivers of employees' engagement (Pitt-Catsouphes and Matz-Costa, 2009). Work engagement is assumed, implicitly, to be gender neutral (Banihani et al., 2013). This gender neutrality assumption indicates that variations in work engagement are down to individual differences and that both women and men can equally demonstrate engagement. Gender is viewed in this study as the social differences between women and men as opposed to just sex (biological differences). It is a system which derives its meaning from an institutionalised system of social practice rather than an individual property. Ely and Padavic (2007, p. 1128) state that, like other systems of differences such as class and race, gender appears in mutually reinforcing aspects such as "resource distribution in societies, hierarchical structures and work practices in organisations, task allocation in families, patterns of interaction between people, and meanings and identities people enact as individuals".

Banihani et al. (2013) argue that work engagement is gendered in that it is easier for men to be engaged that women. Accordingly, men and their characteristics are valued and regarded as more useful than women and their characteristics, leading to a conclusion that processes, practices and interactions in organisations are designed so that it is easier for men to experience psychological meaningfulness, which ultimately leads to more work engagement (Banihani et al., 2013, p. 412). Men are rewarded for emphasising their masculine characteristics and their differences from women, whereas women are usually penalised for their characteristics and differences from men (Williams, 1993). A conclusion that Banihani et al. (2013, p. 414) drew is that men can easily experience psychological safety and be engaged than women in the workplace.

However, Crompton, Lewis and Lyonette (2007) indicate that men and women's experiences in the workplace and at home may influence women's capacity to be fully available and engaged in work and men's capacity to be available and engaged at home. In this note, because the majority of individuals who are responsible for taking care of their families are women, their availability for work may be limited and can be construed as lack of work engagement. Banihani et al. (2013, p. 415) argue that the current presentation of work engagement in the literature and the view that work engagement is the cure of most organisational problems is problematic and gendered. They believe that women have fewer opportunities to experience psychological meaningfulness, safety and availability than men and therefore the notion of work engagement is gendered. Research by Schaufeli et al. (2006) across a number of countries shows inconsistent results in the relationship between work engagement and gender. In the South African sample, women were found to have a higher work engagement than men. Thus, it is assumed in this study that:

H1: Women have higher levels of work engagement than men in a target South African university

Although the new organisational logics appear to be gender neutral on the surface, they remain gendered (Williams, Muller and Kilanski, 2012) and this influences individuals' abilities to demonstrate work engagement. Given that the gendered nature of work engagement remains relatively under-explored in the academic and practitioner literature, Banihani et al. (2013, p. 416) emphasise the need to conduct further empirical studies to explore the nature and extent to which work engagement remains gendered and its implications for female and male employees. To date, no evidence of an existing study which examined the gender influence on work engagement has been found in South Africa and in the South African university context, thus, setting the stage for this empirical investigation. Consequently, this study

seeks to empirically examine the influence of gender on employees' work engagement in a South African university.

4 Methodology

4.1 Research approach

This study followed a quantitative non-experimental cross sectional survey design. Primary data were collected from employees working in a South African university in order to achieve the purpose of the study.

4.2 Research participants

Participants in this study were 154 employees purposively sampled from a South African university. Participants were academic (80%) and support professionals (20%) who had some background in economic and management sciences disciplines. The sample was predominantly African (52%) with the remaining percentage split between Whites (37%), Asians (8%) and Coloureds (3%). Females comprised 61% of the participants relative to 39% being males. About 66% of the participants were married and 54% were in the employment of the institution for at least 5 years. About 59% of the participants had been in their current position for five years.

4.3 Measures

Work engagement was measured using the Utrecht Work Engagement Scale (UWES) (Schaufeli et al., 2002). The UWES is a 17 items self-report instrument that includes three sub-scales, namely, vigour (six items), dedication (five items) and absorption (six items). Responses were measured using a seven-point Likert scale ranging from (1) 'Never' to (7) 'Always'. Sample items include "At my work, I feel bursting with energy", "I am enthusiastic about my job", and "When I am working, I forget everything else around me". The UWES is the most widely used instrument to measure work engagement. A search on PsychINFO showed that 83% of scholarly articles about engagement used this questionnaire (Schaufeli and Salanova, 2011). The structural validity of this instrument has been confirmed in different studies (Salanova, Agut and Peiró, 2005; Llorens, Schaufeli, Bakker and Salanova, 2007; Salanova and Llorens, 2009; Seppälä et al., 2009; Llorens, Bakker, Schaufeli and Salanova, 2006; Salanova, Llorens and Schaufeli, 2011).

4.4 Research procedure

Permission and ethics clearance to conduct the study was obtained from the target university before data collection commenced. The research questionnaire was distributed to the target participants using the "Staff Contact Directory" which lists all staff members employed in the economic and management sciences disciplines. A drop-in and pick-up method was applied to distribute the questionnaires. Questionnaires contained a covering letter which explained the purpose of the study. Participants' consents were obtained prior to data collection and they were advised of the voluntary nature of the research including issues pertaining to anonymity and confidentiality.

4.5 Data analysis

Data were analyzed using the Statistical Package for Social Sciences (SPSS) software (version 23.0). Both descriptive and inferential statistics were computed in line with the purpose of the study. Frequency distributions were computed in order to profile the respondents in terms of their demographic characteristics. After a normality assumption was not held regarding the data, a non-parametric test was conducted to establish the mean differences between males and females with respect to their levels of work engagement. Furthermore, an inter-correlation and multiple regression analyses were computed to establish the correlation matrix and relationship between gender and the three sub-scales of UWES, that is, vigour, dedication and absorption.

5 Results

The researcher first examined the psychometric properties of the UWES which included the following:

- the Kaiser-Meyer-Olkin (KMO) analysis to determine the sample adequacy and the sphericity of the item-correlation matrix.
- the reliability analysis, using Cronbach Alpha coefficients, to measure the accuracy of the UWES and to determine how repeatable the results are.

5.1 Sample adequacy test, item inter-correlations and reliability analysis

The KMO analysis results are depicted in Table 1. The values vary between 0 and 1, and values closer to 1 are better. The suggested minimum value that is acceptable for further analysis is .60 (Tabachnick and Fidell, 2001). The Kaiser-Meyer-Olkin (KMO) index of .942 in this study indicates that the sample is adequate for further statistical analysis. KMO is a measure of how much the items have in common. A KMO value closer to 1 indicates that the variables have a lot in common. The Bartlett's Test of Sphericity which tested inter-item correlation matrix in this study was also conducted. An identity matrix is a matrix in which all the diagonal elements are 1 and off-diagonal elements are 0. The Bartlett's Test of Sphericity was statistically significant (*Df*. 6328; $p \leq .000$), thus supporting the inter-correlations of items of UWES.

Table 1. KMO and Bartlett's Test of Sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.942
Bartlett's Test of Sphericity	Approx. Chi-Square	1816.158
	Df	136
	Sig.	.000

The results of the internal consistency for the UWES and its sub-scales are presented in Table 2. Reliability reflects the consistency of items over time, tests and groups (Kline, 2005; Nunnally and Bernstein, 1994). The UWES scale consists of 17 items in total which were dispersed in three sub-scales (Vigour, Dedication and Absorption). As is evident in Table 2, Cronbach's Alpha coefficient for the UWES is .95 which is very excellent, while that of its scales range from .85 to .88. Generally, Cronbach's alpha ≥ .70 is considered acceptable (Kline, 2005; Polit & Beck, 2004). A reliability coefficient of .70 marks a threshold evidencing high degree of internal consistency (Nunnally, 1978). All sub-scales of the UWES achieved an excellent reliability coefficient, which was considered adequate to continue with further statistical analysis. Thus, the UWES and its sub-scales were found to be very reliable in the current study.

Table 2. Summary of the psychometric properties of the Work Engagement scale and its sub-scales

Scale and sub-scale	Number of items	M	SD	Cronbach's Alpha
Vigour	6	4.36	1.533	.85
Dedication	5	4.50	1.173	.88
Absorption	6	4.30	1.250	.87
Work engagement	**17**	**4.37**	**1.194**	**.95**

5.2 Inter-correlations between gender and UWES sub-scales

Correlations between gender and the sub-scales of UWES were computed and the results are shown in Table 3. It is clear from Table 3 that the inter-correlations among the variables were found to be within the acceptable range because none is ≥ .85 (Bollen, 1989; Almost, 2010) or ≥ .9 (Maiyaki, 2012). Therefore, this is an indication of the absence of multicolinearity problems among the constructs under investigation. Multicolinearity is a problem that occurs when the exogenous variables are highly correlated to as high as .9 and above (Tabachnick and Fidell, 2007). When two or more variables are highly correlated, it means that they contain redundant information and therefore, not all of them are needed in the same analysis (Maiyaki, 2012).

Table 3. Correlations between gender and the UWES sub-scales

	M	SD	1	2	3
1. Gender	1.60	.491			
2. Vigour	26.1513	5.27742	.064		
3. Dedication	22.5033	4.84785	.031	.869**	
4. Absorption	25.7961	5.89908	.124	.871**	.839**

Note: **Correlation is significant at the 0.01 level (2-tailed)

In this study, the following parameters were used to determine the practical effect size of the inter-correlation coefficient values as suggested by Cohen (1988; 1992) and supported by Osteen and Bright (2012): coefficient values around .10 or below were considered small; those around .30 were considered moderate; and those around .50 were considered large in terms of practical significance.

The results in Table 3 show that vigour relates significantly and positively to dedication (r = .869; $p \leq .01$, large practical effect size) and absorption (r = .871; $p \leq .01$, large practical effect size). Dedication also shows a positive and statistically significant relationship with absorption (r = .839; $p \leq .01$, large practical effect size). No statistically significant relationship was established between gender and all three sub-scales of UWES.

5.3 Distribution of normality and test of significant mean differences

The UWES sub-scales were examined for their normality distribution. The Kolmogorov-Sminov Z test was applied and the results are presented in Table 4. Prior to normality distribution testing, the assumption is that data were sampled from a normal distribution or at least from a distribution which is sufficiently close to a normal distribution (Zvi, Turel

and Zerom, 2008). The Kolmogorov-Smirnov test compares the cumulative distribution of the data with the expected cumulative normal distribution, and bases its *p* value ($p \leq .05$) on the largest discrepancy (Öztuna, Elhan and Tüccar, 2006). When normality and homogeneity of variance assumptions are not satisfied, the equivalent non-parametric test must be applied to test mean differences. The results in Table 3 were significant ($p \leq .001$) for all the sub-scales of the UWES.

Table 4. One sample Kolmogorov-Smirnov test

		Normal Parameters[a,b]		Most Extreme Differences			Kolmogorov-Smirnov Z	Asymp. Sig. (2-tailed)
	N	Mean	Std. Deviation	Absolute	Positive	Negative		
Vigour	152	26.1513	5.27742	.103	.066	-.103	.103	.000
Dedication	153	22.5033	4.84785	.106	.061	-.106	.106	.000
Absorption	152	25.7961	5.89908	.098	.048	-.098	.098	.000

Note: a. Test distribution is Normal.
b. Calculated from data.

At the level of significance depicted in Table 4 for all the sub-scales of the UWES, the normality assumption was not held. The null hypotheses that 'the distribution of the sub-scales of the UWES is normal' were rejected. Based on these results, it can be concluded that normality assumptions were untenable and the non-parametric data analyses were justifiable. Consequently, a Mann-Whitney test was conducted to establish if there is a significant mean difference between male and female respondents in relation to their levels of work engagement and the results are depicted in Table 5.

Table 5. Mann-Whitney Test and Mean and Standard Deviation: Gender on UWES sub-scales (N = 154)

Sub-scale	Male			Female			Total			Mann-Whitney Test[a]			
	Mean	*N*	Mean Rank	Mean	*n*	Std. Deviation	Mean	*n*	Std. Deviation	Mann-Whitney U	Wilcoxon W	*Z*	Asymp. Sig. (2-tailed)
Vigour	73.91	59	4360.50	78.15	93	7267.50	26.1513	152	5.27742	2590.500	4360.500	-.580	.562
Dedication	75.96	60	4557.50	77.67	93	7223.50	22.5033	153	4.84785	2727.500	4557.500	-.234	.815
Absorption	70.92	59	4184.00	80.04	93	7444.00	25.7961	152	5.89908	2414.000	4184.000	-1.224	.212

Note: a. Grouping Variable: Gender

It is clear in Table 5 that there are no statistically significant mean differences between male and female respondents in relation to their levels of work engagement. The results show that male and female employees demonstrate almost equal levels of engagement to their work activities.

5.4 Multiple regression analysis

In order to investigate the relationship between gender and the sub-scales of UWES, a multiple regression analysis was undertaken. Such an analysis however assumes variables to have been measured on interval, ratio or dichotomous scales (Tabachnick and Fidell, 2001). The purpose was to test as to whether or not gender does predict respondents' levels of work engagement and the results are depicted in Table 6. Gender was presented as an independent variable whereas the three sub-scales of UWES (Vigour, Dedication and Absorption) were presented as dependent variables.

The regression model depicted in Table 6 explained no presence of variance (Vigour; $R^2 = .004$; Dedication, $R^2 = .001$; and Absorption, $R^2 = .015$ (no practical effect size)) in the dependent variables. In relation to the variance, the multiple regression coefficient was found to be non-significantly different from zero in all three dependent variables (Vigour, $R = .064$, adjusted $R^2 = -.003$ (no practical effect size), $F(.616) = 17.214$; $p \geq .05$; Dedication, $R = .031$, adjusted $R^2 = -.006$ (no practical effect size), $F(.145) = 3.437$; $p \geq .05$; and Absorption, $R = .124$, adjusted $R^2 = .009$ (no practical effect size), $F(2.339) = 80.680$; $p \geq .05$). Furthermore, the results show that gender does not explain the variance in vigour ($\beta = .064$; $p \geq .05$), dedication ($\beta = .031$; $p \geq .05$), and absorption $\beta = .124$; $p \geq .05$). These results indicate that gender does not predict the work engagement of employees at a target university in South Africa.

6 Conclusion

Employee work engagement remains one of the most important aspects of organisational theory that employers should be concerned about globally. This is because engaged employees deliver positive outcomes in the organisation. By implication, an employer's interest on employees' work engagement redirects the energy, involvement and effectiveness that employees bring to the job. The focus of this study was on the examination of gender influence on the levels of work engagement of employees at a South African

university. The findings show that there is no statistically significant relationship between gender and all three sub-scales of Utrecht Work Engagement Scale (that is, vigour, dedication and absorption). No significant mean differences were found between male and female employees in relation to the three sub-scales of UWES.

Table 6. Multiple regression analysis for gender and the three sub-scales of UWES

	Model summary			
	Multiple R	**R^2**	**Adjusted R^2**	**Std. Error of the Estimate**
Standardized data[a, b, c]	.064[a]	.004	-.003	5.28413
	.031[b]	.001	-.006	4.86153
	.124[c]	.015	.009	5.87310

	ANOVA[a, b, c]				
	Sum of Squares	***Df***	**Mean Square**	***F***	**Sig.**
Regression[a, b, c]	17.214[a]	1	17.214	.616	.434[d]
	3.437[b]	1	3.437	.145	.703[d]
	80.680[c]	1	80.680	2.339	.128[b]
Residual[a, b, c]	4188.306[a]	150	27.922		
	3568.811[b]	151	23.635		
	5173.997[c]	150	34.493		
Total[a, b, c]	4205.520[a]	151			
	3572.248[b]	152			
	5254.678[c]	151			

	Coefficients[a, b, c]				
	Unstandardized Coefficients		**Standardized Coefficients**		
	b	***Std. Error***	***β***	***t***	***p***
(Constant)[a, b, c]	25.038[a]	1.481		16.907	.000
	22.010[b]	1.353		16.271	.000
	23.386[c]	1.646		14.208	.000
Gender[a, b, c]	.691[a]	.879	.064	.785	.434
	.307[b]	.805	.031	.381	.703
	1.495[c]	.978	.124	1.529	.128

$R^2 \geq .26$ (large practical effect size); $R^2 \geq .13 \leq .26$ (moderate practical effect size); $R^2 \geq .02 \leq .13$ (small practical effect size)

a. Dependent Variable: Vigour
b. Dependent Variable: Dedication
c. Dependent Variable: Absorption
d. Predictor: Gender

Furthermore, the findings show no presence of variance between male and female employees in relations to the three subscales of UWES. Thus, gender was found not to predict the level of work engagement of employees in a South African university where this study was conducted. These findings are consistent with Schaufeli et al. (2006) finding that there were no gender differences in the level of work engagement for the Canadian, Australian and French samples. However, the current findings contradict Schaufeli et al. (2006) finding that South African and Spanish women showed higher levels of work engagement than men. Banihani et al. (2013) assertion that it is easier for men to be engaged than women is also not supported by the findings of this study. Equally, Thome's (2013) claim that men are likely to be engaged in the workplace than women is also not supported by the findings of this study.

The findings of this study also do not support Kong's (2009) finding that male employees scored higher than females on vigour and absorption while females scored higher than men on dedication. Furthermore, Williams et al. (2012) suggestion that organisations remain gendered to an extent that individuals' abilities to demonstrate work engagement get influenced is also not supported by the findings of this study. Accordingly, men and women in the target South African university have not showed differences with regards to their levels of work engagement. A positive and statistically significant relationship was found between dedication and absorption, meaning that, employees' positive and affective response to their ongoing interactions with significant elements of their jobs and work environment increase their levels of enthusiasm, pride, focus and absorption in their work which make it difficult for them to be detached from such work activities.

A conclusion that can be drawn from the findings of this study is that both men and women at the target South African university have demonstrated

no difference in their levels of work engagement, thus rejecting the notion of gendered work engagement in this particular workplace. A number of factors may be contributing to this, but the most notable is the robust legislative framework (Employment Equity Act 55 of 1998) which sought to create equal opportunities and prohibit any discriminatory practices in the South African workplace. All employees are to be treated equally across all spheres of work irrespective of gender and/or other attributes. All employees in South Africa have a legal recourse and can approach courts of law in the event where they perceive workplace discrimination. The level of literacy of participants in this study who are employees of a university might also be contributing to their understanding of their rights and the obligations imposed by law onto the employer. About 80% of the participants were academic employees and the remaining 20% were managers and support professionals.

Another possible contributing factor is the strength and militant stance of the trade unions which are very active in most sectors in South Africa. Because of the past racial segregation policies which polarized communities in South Africa, most workers have put their trust on trade unions to fight their battles collectively, hence the strength of trade unions. Trade unions are perceived to amass sufficient power to challenge employers in the event of unfair labour practices, and they also act as watchdogs to ensure proper implementation of government policies in the workplace.

Like other studies, this study has some limitations. Firstly, data was collected from employees in one South African university and this makes it improbable to extrapolate the findings of this study to the entire higher education sector or any other sector in South Africa. Secondly, this study focused on gender and its influence on work engagement, not on other demographic variables. Therefore, the interpretation of its findings should be limited only to gender influence in line with the purpose of the study.

Further empirical research is needed to explore the direct influence of employees' gender on their levels of work engagement in other different workplaces/sectors in South Africa and beyond. The notion of gender neutrality on the levels of employees' work engagement in the workplace as supported by the findings of this study prompts further scientific scrutiny into this relationship. A detailed study on ethnic and cultural connotations regarding gender may shed some light on the deeper meaning and interpretation that people attach to the construct of work engagement in a highly diverse South African workplace.

References

1. Almost, J. M. (2010), *Antecedents and consequences of intragroup conflict among nurses in acute care settings* (Unpublished Doctorate Thesis), Canada, University of Toronto.
2. Bakker, A.B. (2009), "Building engagement in the workplace", in Burke, R.J. and Cooper, C.L. (Eds), *The Peak Performing Organization*, Routledge, Oxon, pp. 50-72.
3. Bakker, A.B. and Demerouti, E. (2007), "The job demands-resources model: state of the art", *Journal of Managerial Psychology*, Vol. 22, pp. 309-328.
4. Bakker, A.B, Demerouti, E. and Verbeke, W. (2004), "Using the Job Demands-Resources model to predict burnout and performance", *Human Resource Management*, Vol. 43, pp. 83-104.
5. Bakker, A.B, Gierveld, J.H. and Van Rijswijk, K. (2006), *Success factors among female school principals in primary teaching: A study on burnout, work engagement, and performance*, Right Management Consultants, Diemen, The Netherlands.
6. Bakker, A.B, Schaufeli, W, Leiter, M. and Taris, T. (2008), "Work engagement: an emerging concept in occupational health psychology", *Work & Stress*, Vol. 22, pp. 187-200.
7. Banihani, M, Lewis, P, and Syed, J. (2013), "Is work engagement gendered?", *Gender in Management: An International Journal*, Vol. 28, No. 7, pp. 400 – 423.
8. Barkhuizen, W. (2014), "Spice up employee engagement in your office", *Management Dynamics*, pp. 4-5 (November).
9. Bhatnagar, J. (2012), "Management of innovation: role of psychological empowerment, work engagement and turnover intention in the Indian context", *The International Journal of Human Resource Management*, Vol. 23, pp. 921-951.
10. Bion, W.R. (1961), *Experiences in Groups*, Basic Books, New York, NY.
11. Bollen, K. A. (1989), *Structural equations with latent variables*, John Wiley & Sons, Inc, New York.
12. Burke, R.J. and El-Kot, G. (2010), "Work engagement among managers and professionals in Egypt: Potential antecedents and consequences", *African Journal of Economics and Management Studies*, Vol. 1, No. 1, pp. 42−60.
13. Calas, M. and Smircich, L. (1992), "Rewriting gender into organizational theorizing: directions from feminist perspectives", in Reed, M. and Hughes, M. (Eds), Rethinking Organization: New Directions in Organizational Theory and Analysis, Sage, London, pp. 227-253.
14. Christian, M. and Slaughter, J. (2007), "Work engagement: a meta-analytic review and directions for research in an emerging area", *Academy of Management*, pp. 1-6.
15. Christian, M, Garza, A. and Slaughter, J. (2011), "Work engagement: a quantitative review and test of its relations with task and contextual performance", *Personnel Psychology*, Vol. 64, pp. 89-136.
16. Cohen, J. (1992), "Statistical power analysis". *Current Directions in Psychological Science*, Vol 1, No. 3, pp. 98-101.
17. Cohen, J. (1988), *Statistical power analysis for the behavioural sciences* (2nd ed.), Lawrence Erlbaum, New Jersey.
18. Crompton, R, Lewis, S. and Lyonette, C. (2007), *Women, Men, Work and Family in Europe*, Palgrave Macmillan, London.
19. Ely, R. and Padavic, I. (2007), "A feminist analysis of organizational research on sex differences", *Academy of Management Review*, Vol. 32, pp. 1121-1144.

20. Freud, S. (1922), *Group Psychology and the Analysis of the Ego*, International Psychoanalytic Press, London.
21. Goffman, E. (1961), *Asylums*, Doubleday Anchor, New York, NY.
22. Jeung, C.W. (2011), "The concept of employee engagement: a comprehensive review from a positive organizational behavior perspective", *Performance Improvement Quarterly*, Vol. 24, pp. 49-69.
23. Kahn, W.A. (1990), "Psychological conditions of personal engagement and disengagement at work", *Academy of Management Journal*, Vol. 33, pp. 692-724.
24. Keeffe, P. (2010), *Skills Shortage Demands a Workplace Re-think*, Retrieved April 28, 2015, from Business Managment Suite 101: http://philip-keeffe.suite101.com/skillsshortage-demands-a-workplace-re-think-a246103
25. Kim, H.J, Shin, K.H. and Swanger, N. (2009), "Burnout and engagement: a comparative analysis using the big five personality dimensions", *International Journal of Hospitality Management*, Vol. 28, pp. 96-104.
26. Kline, R. B. (2005), *Principles and practice of structural equation modelling* (2nd ed.), Guilford Press, New York.
27. Kong, Y. (2009), "A study on the job engagement of company employees", *International Journal of Psychological Studies*, Vol. 1, No. 2, pp. 65-68.
28. Koyuncu, M, Burke, R.J. and Fiksenbaum, L. (2006), "Work engagement among women managers and professionals in a Turkish bank: potential antecedents and consequences", *Equal Opportunities International*, Vol. 25, pp. 299-310.
29. Lawler, E.E. (2008), *Talent: Making people your competitive advantage*, Jossey-Bass, San Francisco, CA.
30. Llorens, S, Bakker, A.B, Schaufeli, W.B. and Salanova, M. (2006), "Testing the robustness of the Job Demands – Resource Model", *International Journal of Stress Management*, Vol. 13, pp. 378-391.
31. Llorens, S, Schaufeli, W.B, Bakker, A.B. and Salanova, M. (2007), "Does a positive gain spiral of resources, efficacy beliefs and engagement exists?", *Computer in Human Behavior*, Vol. 23, pp. 825-841.
32. Maiyaki, A. A. (2012), "Influence of service quality, corporate image and perceived value on customer behavioural response: CFA and measurement model", *International Journal of Academic Research in Business and Social Sciences*, Vol. 2, No. 1, pp. 403-414.
33. Martin, J. (2000), "Hidden gendered assumptions in mainstream organizational theory and research", *Journal of Management Inquiry*, Vol. 9, pp. 207-216.
34. Mauno, S, Kinnunen, U. and Ruokolainen, M. (2007), "Job demands and resources as antecedents of work engagement: a longitudinal study", *Journal of Vocational Behavior*, Vol. 70, pp. 149-171.
35. May, D.R, Gilson, R.L. and Harter, L.M. (2004), "The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work", *Journal of Occupational and Organizational Psychology*, Vol. 77, pp. 11-37.
36. Merton, R.K. (1957), *Social Theory and Social Structure*, Free Press of Glencoe, New York, NY.
37. Neumark, D, Johnson, H. and Mejia, M.C. (2011), June. *Future Skill Shortages in the U.S. Economy?*, Retrieved April 28, 2015, from http://www.employmentpolicy.org/sites/www.employmentpolicy.org/files/field-content-file/pdf/David%20Neumark/Neumark%20et%20al%20Skill%20Shortages%20EPRN.pdf
38. Nunnally, J. C. (1978), *Psychometric theory* (2nd ed.), McGraw-Hill, New York, NY.
39. Nunnally, J. C. And Bernstein, I. (1994), *Psychometric theory* (3rd ed.), McGraw Hill, New York.
40. Oberholster, T. (2014), "Rules of engagement", *People Dynamics*, pp. 4-5 (November).
41. Osteen, P. and Bright, C. (2012), *Effect size and intervention research*, University of Maryland, Baltimore.
42. Öztuna, D, Elhan, A. H. and Tüccar, E. (2006), "Investigation of four different normality tests in terms of Type 1 error rate and power under different distributions", *Turkish Journal of Medical Science*, Vol. 36, No. 3, pp. 171-176.
43. Pitt-Catsouphes, M. and Matz-Costa, C. (2009), *Engaging the 21st Century multigenerational workforce: Findings from the age and generations study*, Metlife Mature Market Institute Report. Retrieved 28 April 2015 on http://bc.edu/agingandwork
44. Polit, D. F. and Beck, C. T. (2004), *Nursing research: Principles and methods* (7th ed.), Williams & Wilkins, Philadelphia, Lippincott.
45. Rich, B.L, Lepine, J.A. and Crawford, E.R. (2010), "Job engagement: Antecedents and effects on job performance", *Academy of Management Journal*, Vol. 53, pp. 617-635.
46. Rothbard, N.P. (2001), "Enriching or depleting? The dynamics of engagement in work and family roles", *Administrative Science Quarterly*, Vol. 46, pp. 655-684.
47. Saks, A.M. (2006), "Antecedents and consequences of employee engagement", *Journal of Managerial Psychology*, Vol. 21, pp. 600-619.
48. Saks, A.M. (2008), "The meaning and bleeding of employee engagement: how muddy is the water?", *Industrial & Organizational Psychology*, Vol. 1, pp. 40-43.
49. Salanova, M, Agut, S. And Pieró, J.M. (2005), "Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate", *Journal of Applied Psychology*, Vol. 90, pp. 1217-1227.
50. Salanova, M. and Llorens, S. (2009), "Exposure to information and communication technology and its relationship to work engagement", *Ciencia & Trabajo*, Vol. 32, pp. 55-62.
51. Salanova, M, Llorens, S. and Schaufeli, W.B. (2011), "Yes, I can, I feel good, and I just do it! On gain cycles and spirals of efficacy belief, affect and engagement", *Applied Psychology: An International Review*, Vol. 60, pp. 255-285.
52. Salanova, M. and Schaufeli, W.B. (2008), "A cross-national study of work engagement as a mediator between job resources and proactive behaviour", *The International Journal of Human Resource Management*, Vol. 19, pp. 116-131.
53. Schaufeli, W. and Bakker, A.B. (2004), "Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study",

Journal of Organizational Behavior, Vol. 25, pp. 293-315.
54. Schaufeli, W.B, Bakker, A.B. and Van Rhenen, W. (2009), "How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism", *Journal of Organizational Behavior*, Vol. 30, pp. 893-917.
55. Schaufeli, W.B, and Salanova, M. (2011), "Work engagement: On how to better catch a slippery concept", *European Journal of Work and Organizational Psychology*, Vol. 20, pp. 39-46.
56. Schaufeli, W, Salanova, M, Gonza´lez-Roma, V. and Bakker, A. (2002), "The measurement of engagement and burnout: a two sample confirmatory factor analytic approach", *Journal of Happiness Studies*, Vol. 3, pp. 71-92.
57. Schaufeli, W, Taris, T. and Rhenen, W.V. (2008), "Workaholism, burnout, and work engagement: three of a kind or three different kinds of employee well-being?", *Applied Psychology: An International Review*, Vol. 57, pp. 173-203.
58. Seegers, G. (2012), "Talent crisis threatening business growth and economic prosperity", People Dynamics, Vol. 30, No. 11, p. 21 (November).
59. Seppälä, P, Mauno, S, Feldt, T, Hakanen, J, Kinnunen, U, Tolvanen, A. and Schaufeli, W.B. (2009), "The construct validity of the Utrecht Work Engagement Scale: Multisample and longitudinal evidence", *Journal of Happiness Studies*, Vol. 10, pp. 459-481.
60. Sharp, L. (2011), *South Africa's extraordinary skills shortage – Adcorp*, Retrieved April 28, 2015, from http://www.politicsweb.co.za/politicsweb/view/politicsweb/en/page71619?oid=235542&sn=Detail&pid=71619
61. Simpson, M.R. (2009), "Engagement at work: a review of the literature", *International Journal of Nursing Studies*, Vol. 46, pp. 1012-1024.
62. Slater, P.E. (1966), *Microcosms*, Wiley, New York, NY.
63. Smith, K.K. and Berg, D.N. (1987), *Paradoxes of Group Life*, Jossey-Bass, San Francisco, CA.
64. Tabachnick, B. G. and Fidell, L. S. (2007), *Using multivariate statistics* (5th ed.), Allyn & Bacon, New York.
65. Tabachnick, B. G. and Fidell, L. S. (2001), *Using multivariate statistics* (4th ed.), Allyn & Bacon, Needham Heights, MA.
66. Taris, T, Cox, T. and Tisserand, M. (2008), "Engagement at work: an emerging concept", *Work & Stress*, Vol. 22, pp. 185-186.
67. Thome, E. (2013), *Employee engagement: Why is gender (still) a factor?*, TalentCulture, Retrieved on 28 April 2015 from http://www.talentculture.com.
68. Welbourne, T.M. (2007), "Employee engagement: beyond the fad and into the executive suite", *Leader to Leader*, Vol. 2007, pp. 45-51.
69. Williams, C. (1993), *Doing Women's Work: Men in Non-Traditional Occupations*, Sage, London.
70. Williams, C.L, Muller, C. and Kilanski, K. (2012), "Gendered organizations in the new economy", *Gender & Society*, Vol. 26, pp. 549-573.
71. Xanthopoulou, D, Bakker, A.B, Demerouti, E. and Schaufeli, W.B. (2007), "The role of personal resources in the job demands-resources model", *International Journal of Stress Management*, Vol. 14, pp. 121-141.
72. Zieminski, N. (2009), *Help still wanted, global talent crunch persists*, Retrieved April 28, 2015, from Reuters: http://www.reuters.com/article/2009/05/28/us-manpowertalent-idUSTRE54R0SO20090528
73. Zvi, D, Turel, O. and Zerom, D. (2008), *A modified Kolmogorov-Smirnov test for normality*, Munich Personal RePEc Archive, Munich.

TRADE MISINVOICING, EXTERNAL DEBT AND SUSTAINABLE DEVELOPMENT: A NIGERIAN EXAMPLE

Collins C Ngwakwe*

Abstract

This paper evaluated how trade misinvoicing orchestrates external debt in Nigeria and its obstructive tendencies on Nigeria's sustainable economic development. The paper is pertinent, given that Nigeria is among the top ten developing countries in the world who are victims of substantial illicit trade misinvoicing outflows. The methodological approach is a mix of descriptive analysis (using tables and graphs) and a t-test of difference in means between trade misinvoicing outflow from Nigeria, external debt and official development assistance (OD) in Nigeria for the period 2003 – 2012. Findings indicate that as trade misinvoicing outflow increased during the period 2003 -2012, Nigeria's external debt increased yearly. Results from the statistical t-test showed that the mean difference in trade misinvoicing outflow is significantly greater than the mean differences in external debt and official development assistance received into Nigeria. This finding attests to the huge internal financial resources that Nigeria lost during the period 2003 - 2012 through illicit trade misinvoicing outflow. The analysis further disclosed that trade misinvoicing outflow has hampered Nigeria's stride to sustainable economic development given the record increases in unemployment, poverty, lack of access to sanitation facilities, low percentage of qualified health staff to child birth and a widening income inequality as measured by GINI index. The paper concludes that the drainage of Nigeria's internal financial resources through illicit trade misinvoicing has denied Nigeria the needed finance to enhance the actualisation of sustainable economic development. Recommendations are proffered to assist in halting trade misinvoicing outflow from Nigeria.

Keywords: Trade Misinvoicing, External Debt, Sustainable Development, Illicit Outflow, Taxation, Poverty, Nigeria

**Turfloop Graduate School of Leadership, Faculty of Management and Law, University of Limpopo, South Africa*

1 Introduction

This paper attempts to examine how trade misinvoicing has orchestrated external borrowing in Nigeria and its obstreperous propensities in the achievement of Nigeria's sustainable economic development. The paper is deemed pertinent given that Nigeria and South Africa are the only African countries rated among the top ten developing countries in the world who are victims of substantial illicit trade misinvoicing outflows. It is apparent that the corporate and other informal businesses abuse the national and international trading and tax systems, hence many commercial transactions' real values are advertently obscured through trade misinvoicing to obviate taxation in host countries, which thus render the developing nations poorer, hence draining the necessary resources for poverty eradication, sustainable growth and development (Global Financial Integrity, 2015a; Wu, 2010). There is a growing shocking statistic indicating the extent to which trade mininvoicing schemes, via illicit financial flows, (Global Financial Integrity [GFI], 2015a) shift finances and other assets abroad to avoid tax obligations in developing host countries (Wu, 2010). Surprisingly, this seemingly obscured genre of a corporate profiteering scheme does not seem to receive the expected societal attention and condemnation compared to the prevalent censure of public sector corruption. Fortunately, the Global Financial Integrity (GFI) is doing a sterling job in this regard by blowing the whistle to alert the world about the sustainable development implications of resource deprivation meted out to developing nations through illicit financial flows of which tax revenue loss through trade misinvoicing constitutes an incredibly gargantuan amount, hence the GFI bemoans that of all the illicit financial flows: *"The proceeds of commercial tax evasion, mainly through trade mispricing, are by far the largest component"* (Global Financial Integrity, 2010:1). Accordingly recent official statistics indicate that trade misinvoicing constitutes about 80% of all illicit financial outflows (IFF) from developing nations (Global Financial Integrity, 2013). This heinous amount of loss in government revenue may appear somewhat concealed probably because of foreign direct investment (FDI) inflows to development countries (UNCTAD, 2014); however a disturbing revelation about the FDI inflows to developing countries is that firstly, the amount of

trade misinvoicing outflow and other illicit financial outflow from developing countries dwarf the amount of FDI inflow to these countries. Secondly is the concern raised by Christian Aid (2013) that a greater percentage of the FDI come from tax havens and/or secrecy jurisdictions, where also it is alluded that money siphoned from developing nations through illicit financial flows find its destination – *give little and take more conundrum, or round tripping;* according to Reuters (2013). According to Global Financial Integrity (2014), about 80% of illicit financial outflows from all the developing countries are perpetrated via trade misinvoicing. In Sub-Saharan Africa alone, over 68% of illicit financial outflow is via trade misinvoicing (Global Financial Integrity, 2014:25). These figures far exceed the official development assistance received by developing nations:

"The amount of money that has been drained out of Africa – hundreds of billions decade after decade – is far in excess of the official development assistance going into African countries,"(Global Financial Integrity, 2010: 1)

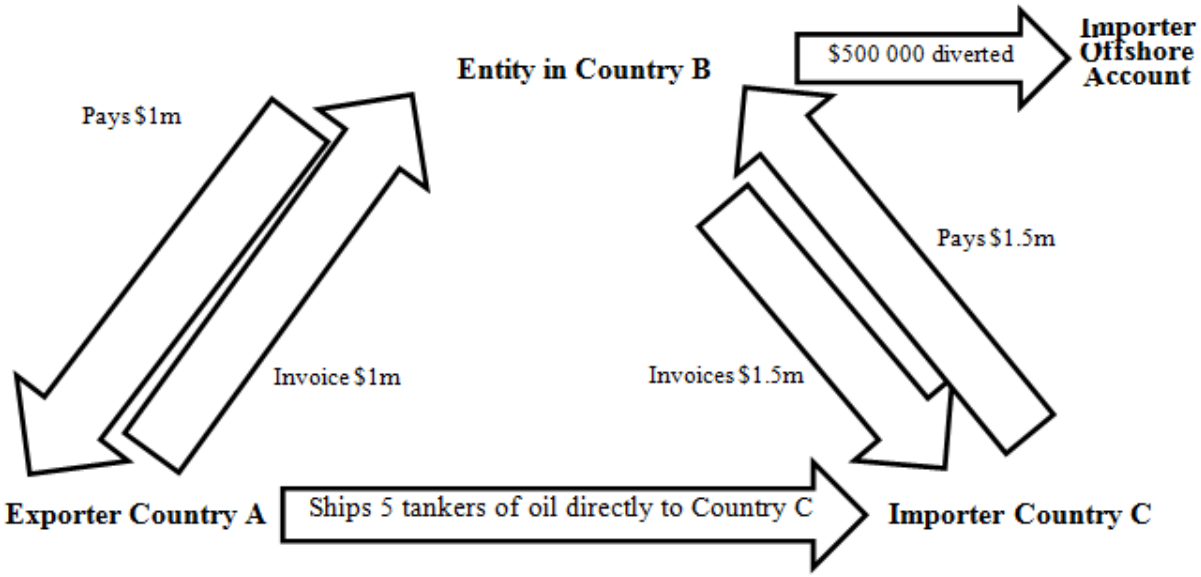
The above revelation seems to infer that if money that is siphoned from developing countries via trade misinvoicing is retained, the rate of poverty and unemployment may decline as developing countries such as Nigeria would have more resources to provide health care, education and infrastructure (Tafirenyika, 2013; OECD, 2014). However, given the staggering amount of funds that Nigeria loses to trade misinvoicing outflow annually, little revenue is left for the government to provide sufficient social and infrastructure services. The alternative thus is heavy reliance on external borrowing. But external debt accumulation with associated cost of capital tends to pose an impediment to sustainable socio-economic development. Accordingly, the question that underpins this paper is how trade misinvoicing outflow compares with the growth of external debt and how this has posed an obstacle to sustainable socio-economic development in Nigeria. Accordingly, the objective of this paper is to discuss and illustrate how trade misinvoicing outflow orchestrates reliance on external debt and the concomitant implication for sustainable socio-economic development in Nigeria.

The rest of this paper is organised as follows; the next section, following the introduction, presents the review of related literature. The section after that presents the method, analysis and discussion. The final section presents the conclusion.

2 Related literature

According to (Buehn & Eichler, 2011, p.1263) *"trade misinvoicing occurs if the true value of exports or imports deviates from the amount of exports or imports businesses report to the authorities".* It is a scheme for siphoning money across international borders by deliberately over or under invoicing the price of commercial transactions, manipulation of the quantity and quality of transactions, such that the invoice submitted to the customs fail to reflect the real price of the transactions (Global Financial Integrity, 2015b). A schematic representation of the trade misinvoicing process is presented in Figure 1.

Figure 1. How Trade Misinvoicing Occurs



Source: adapted from the Global Financial Integrity (2015b, p.3) Trade misinvoicing, available at:http://www.gfintegrity.org/issue/trade-misinvoicing/ [accessed April 10 2015].

The above chart depicts a typical case of import over-invoicing adapted from the Global Financial Integrity (2015b:3), although the importer in country C's real value of imported oil from country A is $1m,

it uses its trade intermediary in country B to re-issue an invoice which increases the amount of import to $1.5m; hence the importer in country C pays $1.5m to its intermediary in country B who in turn settles the country A exporter $1m; the importer in country C subsequently uses its intermediary in country B to send the $500 000 over-invoiced amount to its offshore bank account (see: e.g. Global Financial Integrity, 2015b:3). See also an example of a recent case in the USA about import-export over and under invoicing instances (Buenos Aires Herald, 2014). For more detailed information on import and export misinvoicing see Financial Action Tax Force, (2006). Trade misinvoicing moves money in and out of a country to the benefit of business and to the detriment and denial of tax revenue to the countries affected: "*money is moved out of a country by under-invoicing exports or over-invoicing imports. Money is moved into a country by over-invoicing exports or under-invoicing imports"* (International Trade Alert, 2015:1). Import over-invoicing may be prevalent in developing countries whose customs have adopted import price surveillance to protect local infant industries from unfairly low prices of imported goods:

"One possible impact of the import surveillance mechanism is over invoicing of imports in order to avoid surveillance procedures since the procedures are applied to goods with prices under a predefined reference price"(Atkas and Aldan, 2013: 1)

The importer makes illicit gain in two ways, the importer over-invoices to avoid customs surveillance and the payment of relevant customs import duties; again the importer also uses the over-invoiced import amount to overstate the operating expenses in the annual income statement, leading to understatement of the period income and hence avoids taxation – in turn, the country loses customs revenue and corporate income tax revenue resulting from a single misinvoicing scheme. An aggregate loss of revenue from numerous importers and exporters in a developing country leads to millions of dollars loss in revenue of a developing country.

It has therefore been established that a major conduit for illicit financial outflow from developing countries is trade misinvoicing (Aizenman, 2008; Global Financial Integrity, 2014). Aizenman posits that the greater the commercial openness in an economy, the greater the degree with which trade misinvoicing orchestrates illicit financial outflow (Aizenman, 2008), and sub-Saharan Africa suffers heavy illicit outflow amounting to 5.5% of the region's GDP – making it the highest in terms of illicit outflow to GDP (Global Financial Integrity, 2014). Consequently, in order to finance development projects, the sub-Saharan African countries embark on international borrowings which in turn makes the region a net creditor to international finance institutions (Boyce & Ndikumana, 2001).

Accordingly, the Global Financial Integrity warns that national revenues drained through commercial tax avoidance via trade misinvoicing is of a gargantuan proportion when compared to a small percentage (about 3%) of corruption's contribution to illicit financial outflow (Global Financial Integrity, 2014: 1). Nevertheless the small percentage of public sector corruption cannot be taken for granted as corruption is an incubator of trade misinvoicing schemes. "*Corruption, while only a small share of overall illicit flows, is a cross-cutting driver of such flows and risks undermining any efforts to curb them"* OECD (2014: 2). However some public sector corruption has been seen as being orchestrated by some corporates;

it is alluded that some huge corporate perpetrates trade misinvoicing by weakening the governance processes in some developing countries whose economies are far too low when compared to the wealth of some multinational firms with colossal affluence and influence; hence, with financial enticement, some multinationals wield intimidating control over the governmental processes of developing nations (see e.g. UNECA, 2013). Such weakening thus renders the government's arms vulnerable to corruption which thus opens up the channel for more trade misinvoicing and other forms of illicit financial outflows. The matter is made worse by weak regulations and laws in some developing nations (Amaeshi et al., 2006; Jansky, 2013) – these thus make the perpetrators of trade misinvoicing somewhat invincible. Consequently*"Trade misinvoicing remains by far the most popular way to illicitly move money out of developing countries, comprising 77.8 percent of the global ten-year IFF total in real terms*" (Global Financial Integrity, 2014: 15).

According to Global Financial Integrity (2014), trade misinvoicing accounts for 77.8% (roughly 80%) of illicit financial flows from developing nations and this decimates domestic resources and heightens poverty and thus retards development; this is also emphasised by the Organisation for Economic Corporation and Development (OECD):

"Illicit financial flows strip developing countries of resources that could be used to finance much needed public services, such as health care and education. A mere portion of these funds would have a significant positive impact: it is estimated that every $100 million recovered could fund full immunisations for 4 million children or provide water connections for some 250 000 households in a developing country" (OECD, 2014: 3).

Trade misivoicing outflows from developing countries is therefore a bane to social and economic development because it:

"…drains hard currency reserves, heightens inflation, reduces tax collection, cancels investment, and undermines free trade. It has its greatest impact on those at the bottom of income scales in their countries, removing resources that could otherwise be

used for poverty alleviation and economic growth. (Global Financial Integrity, 2010: 1)

Illicit financial flows reduce the availability of funds to the government and thus widens income inequality (Jansky, 2013:6; UNECA, 2013). It also reduces private funds and contributes to low investment and hence unemployment (Jansky, 2013). Worse still is, according to UNECA (2013), trade misinvoicing and other illicit financial outflow schemes weaken governance. This is because, with their financial girth, huge companies penetrate government entities and with financial enticements (The Fiscal Times, 2011) they create a fissure along the channels of economic and fiscal policy monitoring systems and/or organs, such that these organs tend to turn a 'blind eye' when trade rules are flouted, (The Fiscal Times, 2011; Stephens, 2002). Furthermore, due to the huge investments of multinational entities, they tend to command an intimidating influence on government policies of most developing countries as the wealth of some multinational entities surpass the wealth of some developing countries (Gray & Bebbington, 1998; Sklair, 2002); "*The largest TNCs have assets and annual sales far in excess of the GNP of most of the countries in the world* (Sklair, 2002:36). Thus, advertently or inadvertently, some governments of developing economies kowtow to the seemingly colossal influence of multinationals; "*the sheer size and power of the TNCs place them beyond the control of States - that TNCs are, indeed, no longer controllable" (Gray and Bebbington, 1998:.6)*

The growth of trade misinvoicing is seen to elude control, given that many developing nations have frail legislative and weak institutional frameworks for law enforcement to stem the abuse of trade misinvoicing and other transfer pricing schemes (Jansky, 2013:6), these countries are thus vulnerable to a high loss of tax revenue with concomitant social cost consequence. Given the over 68% illicit financial outflow through trade misinvoicing schemes from Nigeria, efficiency of the regulatory system becomes doubtful (Amaeshi et al., 2006). Whilst it has been recognised that trade misinvoicing is a major conduit for drainage of finance from developing countries (Global Financial Integrity, 2014), corruption is recognised in the literature as an incubator and facilitator of illicit financial outflows orchestrated by companies and other entities (Christensen, 2012), hence policies that are designed to nip trade misinvoicing must also incorporate public sector venality. The following sections present a descriptive and quantitative comparison of trade misinvoicing, external debts and official development assistance in Nigeria.

3 Methodology

In order to visualise the state of and apparent implication of trade misinvoicing outflow from Nigeria, the researcher applied a mix of descriptive analyses approach (visual descriptions using tables and graphs and a quantitative summary of statistics – in this case a t-test of difference in means) in the analysis of data between trade misinvoicing outflow from Nigeria, external debt and official development assistance (OD) in Nigeria for the period 2003 – 2012. Whilst researchers have disagreed about which method of presenting data is better – graphs or tables (Gelman, 2011), they however concur that tables and graphs are useful in research information communication (Gelma et al., 2002). Hence the researcher chose to utilise tables with graphs and charts because graphs communicate the information clearly to a wider audience and thus enhances the accessibility and understanding of research results (Kastellec & Leoni, 2007). In addition to tables and graphs, the researcher also applied the T-test of difference in means to ascertain which mean is greater and at what degree of significance exists between the means of trade misinvoicing outflow and external debt.

4 Comparative Analysis of Illicit Financial Outflow (IFF) and Trade Misinvoicing outflow from Nigeria 2003-2012.

Sub-Saharan Africa represented 8.0 percent of combined illegal financial flows from the developing world for the period 2003-2012, and the two biggest African economies, South Africa and Nigeria, are amongst the top ten developing countries in the world with significant financial outflows through trade misinvoicing:

"There are two Sub-Saharan African countries in the top ten globally: Nigeria and South Africa. IFFs averaged 5.5 percent of the region's GDP over this ten-year period. A significant majority of IFFs from Sub-Saharan Africa – 68.2 percent – were due to trade misinvoicing." (Global Financial Integrity, 2014:25)

This implies that, out of the total illicit financial outflows from Nigeria during 2003 - 2012, more than 68% was through trade misinvoicing; translating this loss into a monetary amount, the total illicit financial outflows from Nigeria during 2003 – 2012 amounted to $157.4bil (Global Financial Integrity, 2014:13) and a whopping fraction of this amount – 68% ($107bil) was through trade misinvoicing (see tables and graphs in subsequent sections). Nevertheless, the GFI maintains that its figures are conservative estimates (Global Financial Integrity, 2014). It is thus likely that trade misinvoicing outflows from Nigeria during the period 2003 - 2012 may have been more than the aforesaid conservative amounts.

Table 2 and Figure 2 present a comparison of illicit financial outflow from Nigeria and trade misinvoicing component.

Table 1. Illicit Financial outflows and Trade Misinvoicing Outflow from Nigeria 2003-2012

	US$M	US$M
	IFF	T.Misinvoicing (68% of IFF)
2003	0	0
2004	1680	1142.4
2005	17867	12149.56
2006	19159	13028.12
2007	19335	13147.8
2008	24192	16450.56
2009	26377	17936.36
2010	20780	14130.4
2011	20144	13697.92
2012	7922	5386.96
	157456	107070.1

Source: author's table with Illicit Financial Flow data from Global Financial Integrity (2014:13); trade misinvoicing is computed by the author based on 68% ratio of total illicit financial flows from Nigeria as indicated by Global Financial Integrity (2014:25).

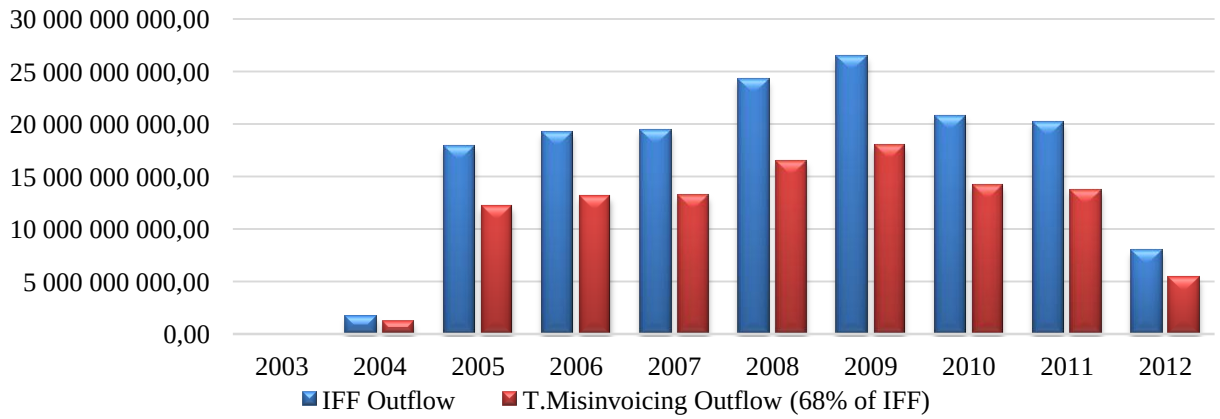
Given that the above amounts were presented in millions by the Global Financial Integrity (2014:13), the author converted these amounts to millions by multiplying each amount by one million (1 000 000), thus the actual amounts of illicit financial flows and associated 68% (Global Financial Integrity 2014:25) trade misinvoicing flows out of Nigeria is presented in Table 2.

Table 2. Comparison between Illicit Financial outflows and Trade Misinvoicing Outflow from Nigeria 2003-2012

Year	IFF Outflow from Nigeria US$	Trade Misinvoicing Outflow from Nigeria (68% of IFF) US$
2003	0.00	0.00
2004	1 680 000 000.00	1 142 400 000.00
2005	17 867 000 000.00	12 149 560 000.00
2006	19 159 000 000.00	13 028 120 000.00
2007	19 335 000 000.00	13 147 800 000.00
2008	24 192 000 000.00	16 450 560 000.00
2009	26 377 000 000.00	17 936 360 000.00
2010	20 780 000 000.00	14 130 400 000.00
2011	20 144 000 000.00	13 697 920 000.00
2012	7 922 000 000.00	5 386 960 000.00
Total 2003-2012	157 456 000 000.00	107 070 080 000.00

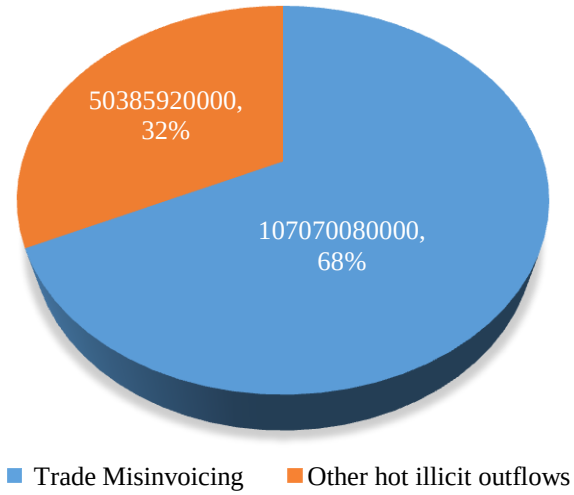
Source: author's table with data from Global Financial Integrity (2014:13)

Figure 2. Bar Chart Comparison between Illicit Financial outflows and Trade Misinvoicing Outflow from Nigeria 2003-2012 US$



Source: author's bar chart with data from table 2

Figure 3. Pie Chart Comparison between Illicit Financial outflows and Trade Misinvoicing Outflow from Nigeria 2003-2012 US$



Source: author's pie chart with data from table 2

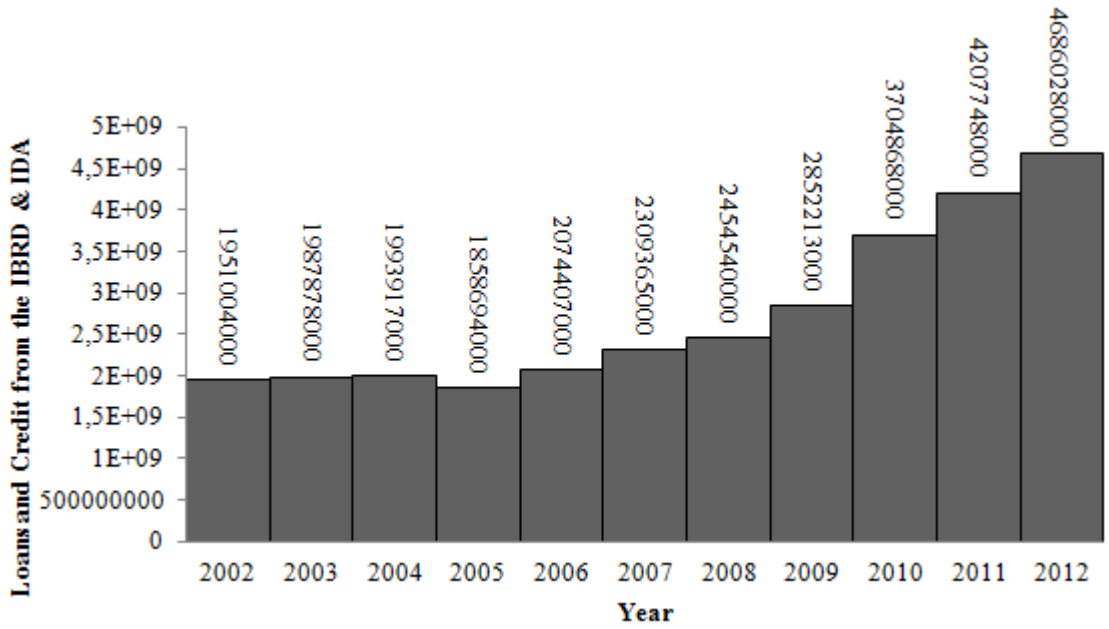
According to UN (2013), many of the developing countries who are victims of substantial trade misinvoicing schemes and other illicit financial outflows suffer from huge external debts. This thus places the affected countries in a dilemma of choice between using the meagre national resources to provide essential services to the citizenry or to service external debts. Irrespective of the choice made though, the opportunity cost of the forgone alternative remains huge – failure to service the external debts increases the cost of debts and failure to provide social services gives rise to a denial of basic social services that thus deepens inequality, poverty and retardation of inclusive growth and development. Being amongst the top ten developing countries that are victims of trade misinvoicing (Global Financial Integrity, 2014:25), during the period 2003 – 2012, Nigeria lost billions of dollars in revenue through trade misinvoicing. Within the same period, Nigeria's external debt grew steadily in billions of dollars. Thus the alternative to raising development revenue is apparently through external borrowing. Table 3 and Figure 4 present a comparative analysis of trade misinvoicing outflow from Nigeria and the growth of external debt.

Table 3. Nigeria's External Debt 2002 – 2012

Year	Nigeria External Debt IBRD/IDA Loan US$
2002	1951004000
2003	1987878000
2004	1993917000
2005	1858694000
2006	2074407000
2007	2309365000
2008	2454540000
2009	2852213000
2010	3704868000
2011	4207748000
2012	4 686 028 000

Source: World Bank (2014) IBRD loans and IDA credits (DOD, current US$) http://data.worldbank.org/indicator/DT.DOD.MWBG.CD

Figure 4. Bar Graph of Nigeria's External Debt 2002 – 2012 US$



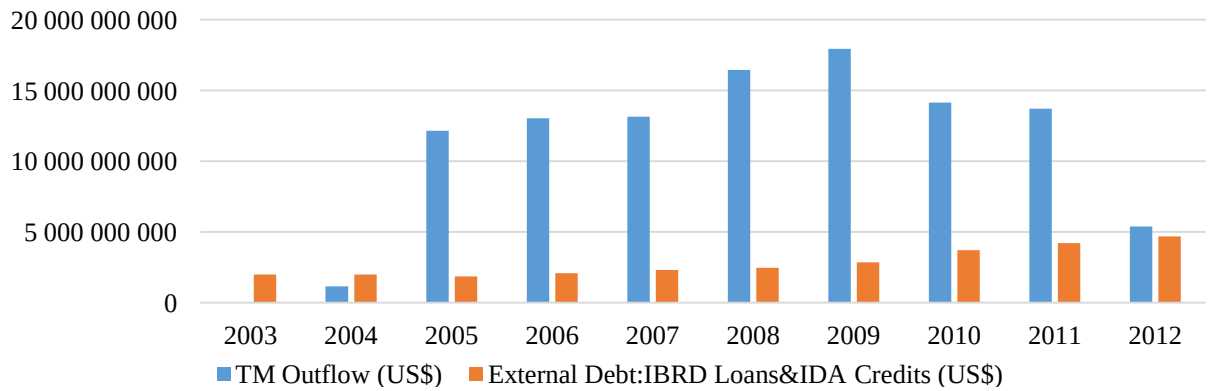
Source: author's graph with data from the World Bank (2014) IBRD loans and IDA credits (DOD, current US$) http://data.worldbank.org/indicator/DT.DOD.MWBG.CD

Table 4. Comparative table of Trade Misinvoicing outflows and External Debt in Nigeria 2003-2012

Year	TM Outflow (US$)	External Debt:IBRD Loans&IDA Credits (US$)
2003	0	1 987 878 000
2004	1 142 400 000	1 993 917 000
2005	12 149 560 000	1 858 694 000
2006	13 028 120 000	2 074 407 000
2007	13 147 800 000	2 309 365 000
2008	16 450 560 000	2 454 540 000
2009	17 936 360 000	2 852 213 000
2010	14 130 400 000	3 704 868 000
2011	13 697 920 000	4 207 748 000
2012	5 386 960 000	4 686 028 000
Total 2003-2012	107 070 080 000	28 129 658 000

Sources: World Bank (2014) IBRD loans and IDA credits (DOD, current US$), available from http://data.worldbank.org/indicator/DT.DOD.MWBG.CD; Global Financial Integrity (2014) illicit financial flows from developing countries 2003-2012 http://www.gfintegrity.org/wp-content/uploads/2014/12/Illicit-Financial-Flows-from-Developing-Countries-2003-2012.pdf

Figure 5. Comparative Graph of Trade Misinvoicing outflows and External Debt in Nigeria 2003-2012



Source: author's bar Graph, with data from Table 4

5 Mean Difference between Trade Misinvoicing Outflow & Nigeria's External Debt 2003–2012

Table 5. t-Test: Paired Two Sample for Means between Trade Misinvoicing Outflow Nigeria's External Debt 2003 -2012

Hypothesis:
$H1_0$: Mean TM < Mean ExtDebt,
$H1_1$: Mean TM > Mean ExtDebt,
tested at alpha:0.05, accept $H1_1$ *if P<0.05*

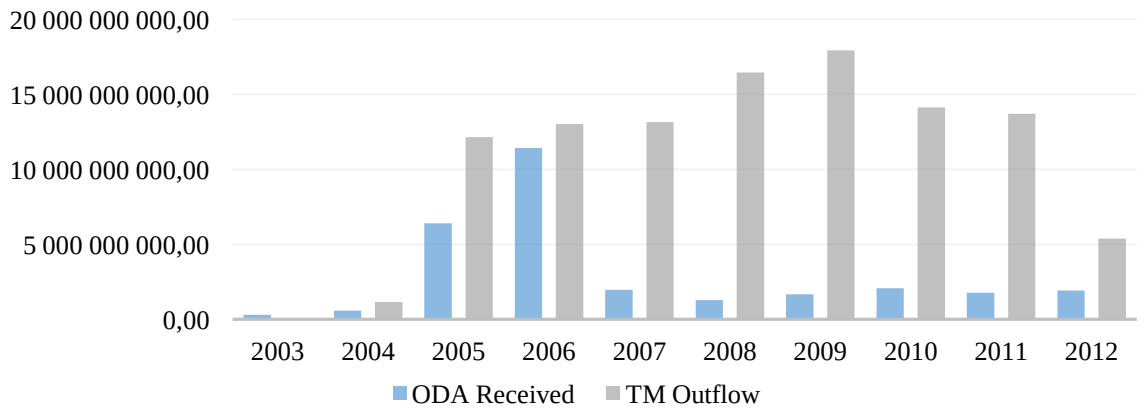
	TM Outflow (US$)	*External Debt:IBRD Loans&IDA Credits (US$)*
Mean	10707008000	2812965800
Variance	3.93075E+19	1.04896E+18
Observations	10	10
Pearson Correlation	0.135546875	
Hypothesized Mean Difference	0	
df	9	
t Stat	4.017143099	
P(T<=t) one-tail	0.001515538	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.003031075	
t Critical two-tail	2.262157163	

Table 6. Official Development Assistance & Official Aids Received (ODA) Compared with Trade Misinvoicing Outflow from Nigeria 2003-2012

Year	ODA Received US$	TM Outflow US$
2003	308 220 000.00	0.00
2004	576 940 000.00	1 142 400 000.00
2005	6 408 810 000.00	12 149 560 000.00
2006	11 428 020 000.00	13 028 120 000.00
2007	1 956 260 000.00	13 147 800 000.00
2008	1 290 160 000.00	16 450 560 000.00
2009	1 657 070 000.00	17 936 360 000.00
2010	2 061 960 000.00	14 130 400 000.00
2011	1 768 550 000.00	13 697 920 000.00
2012	1 915 820 000.00	5 386 960 000.00
Total 2003 - 2012	29 371 810 000.00	107 070 080 000.00

Sources: ODA Received, World Bank (2015a) Net official development assistance and official aid received (current US$), http://data.worldbank.org/indicator/DT.ODA.ALLD.CD

Figure 6. Bar Graph of Nigeria Official Development Assistance Received Compared with Monetary Lost Through Trade Misinvoicing 2003-2012



Source: author's Bar Graph with data from Table 6

Table 7. Paired Two Sample for Means between Trade Misinvoicing Outflow from Nigeria and Official Development Assistance Received

Hypothesis:
$H1_0$: Mean TM < Mean ODA Received
$H1_1$: Mean TM > Mean ODA Received
tested at alpha:0.05, accept $H1_1$ *if P<0.05*

	TM Outflow	*ODA Received*
Mean	10707008000	2937181000
Variance	3.93075E+19	1.16797E+19
Observations	10	10
Pearson Correlation	0.276256308	
Hypothesized Mean Difference	0	
df	9	
t Stat	3.926923606	
P(T<=t) one-tail	0.001737118	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.003474235	
t Critical two-tail	2.262157163	

Furthermore, the Global Financial Integrity laments that illicit financial outflows from developing countries exceed the official development assistance (ODA) received by these nations.

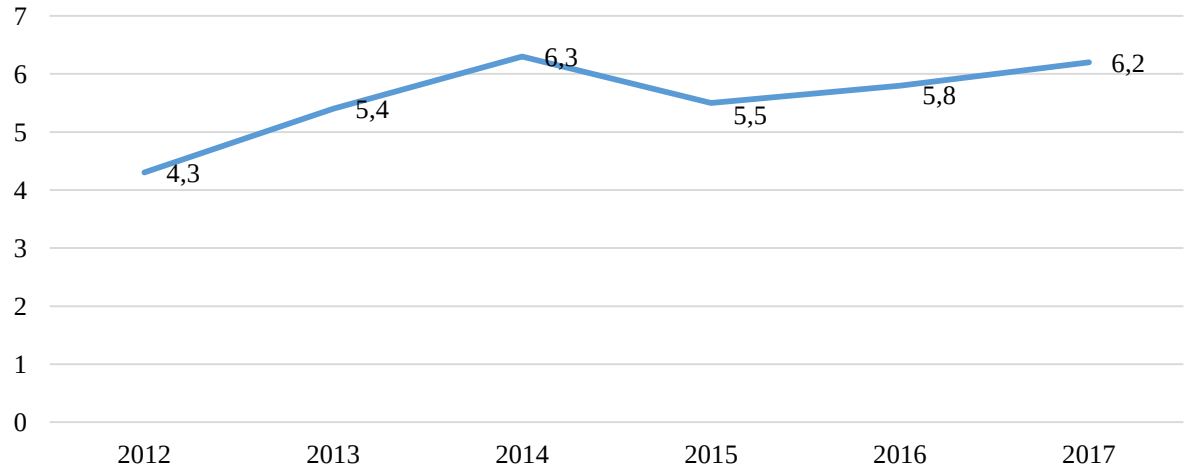
"Illicit outflows were roughly 1.3 times the US$789.4 billion in total FDI, and they were 11.1 times the US$89.7 billion in ODA that these economies received in 2012" (*Global Financial Integrity, 2014a:4)*

This thus implies that trade misinvoicing outflow may equate or be greater than the ODA received by Nigeria during the period 2003 – 2012. This is also presented in Table 6.

6 Implication on Nigeria's Sustainable Development

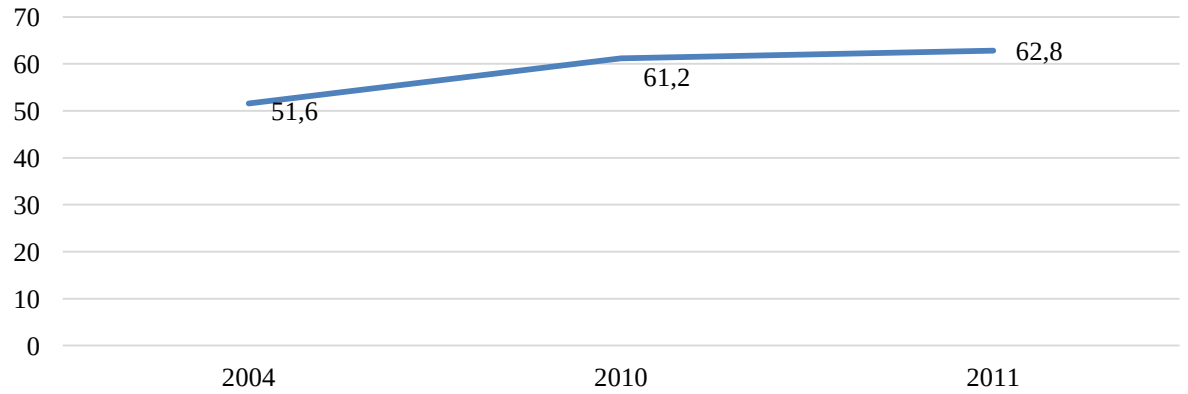
The preceding analyses show that during the period 2002 - 2012 Nigeria experienced huge amounts of revenue drained out of its shores through trade misinvoicing, it is also evident that this period of trade misinvoicing corresponded with the growth in Nigeria's external debt. Given therefore that the amount lost through trade misinvoicing exceeds the amount of external borrowing and also exceeds the amount received as official development assistance (ODA), it becomes glaringly clear that Nigeria lost a huge fund that could have been used to enhance the actualisation of sustainable development in Nigeria through infrastructure development and investment. Hence, this section gives brief attention to the performance of some sustainable development indicators such as poverty and employment levels during this period. Albeit, the fact that Nigeria has had an improved GDP growth (Figure 7), and the growth is estimated to continue along the same trajectory, the GDP growth still does not comprehensively represent how well the wealth or income of a nation is distributed and hence not a balanced test of sustainable social, environmental and economic development. A rising GDP does not necessarily lead to advancement in *"Ecosystem Sustainability', 'Health and Wellness' and 'Opportunity"* (Social Progress Imperative, 2014:1), the GDP is thus seen as a subjective measure of wellbeing – its growth does not necessarily translate to growth in every household's income level and standard of living (Diener et al., 2013). In their research Dierner et al. (2013:274) found that GDP growth resulted in only 18% shift in household income levels. This finding implies that, in some cases, only a small percentage of the population may benefit from a rising GDP of a country. Accordingly, the Organisation for Economic Cooperation and Development is worried that the GDP can be misleading, it obscures the state of well-being of the people, it does not show how a nation's wealth is distributed, and it does little to narrow the gap between the rich and the poor (OECD, 2015). Therefore, a new initiative by the OECD is underway to develop a new and more robust, inclusive indicator of measuring progress and wealth (OECD, 2015). Consequently, in order to visualise the economic and social dent meted out by trade misinvoicing outflow from Nigeria, it is therefore pertinent to look beyond Nigeria's GDP and examine the trend in poverty, employment and equitable distribution of consumption income during these years of trade misinvoicing outflow from Nigeria. These appear in schematic presentations below.

Figure 7. Line Chart of Nigeria's GDP Growth



Source: author's line chart with data from: World Bank (2015b) Country and region specific forecasts and data: GDP and Current Account Balance growth forecasts by country, region, and income level. Available from: http://www.worldbank.org/en/publication/global-economic-prospects/data?region=SST

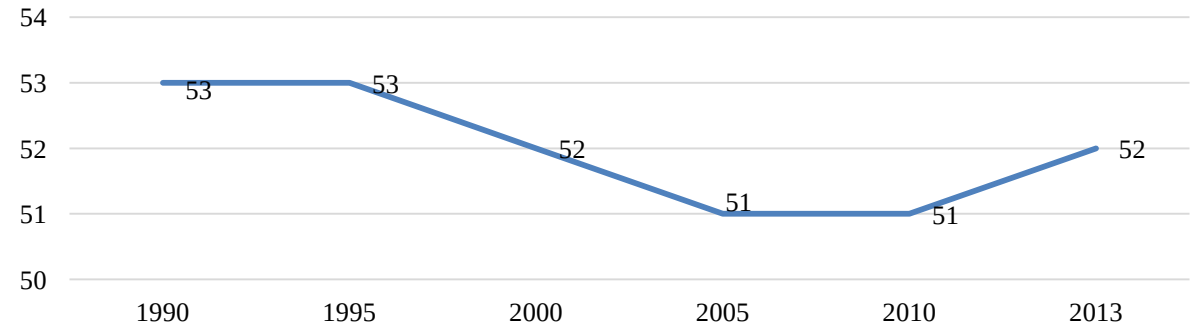
Figure 8. Line Chart Nigeria's Poverty Head Count Ration (PPP) (% of Population below $1 a day)



Source: author's line chart with data from: National Bereau of Statistics Nigeria (NBS) (2012) Nigerian poverty profile 2010 report, available at: http://reliefweb.int/sites/reliefweb.int/files/resources/b410c26c2921c18a6839baebc9b1428fa98fa36a.pdf

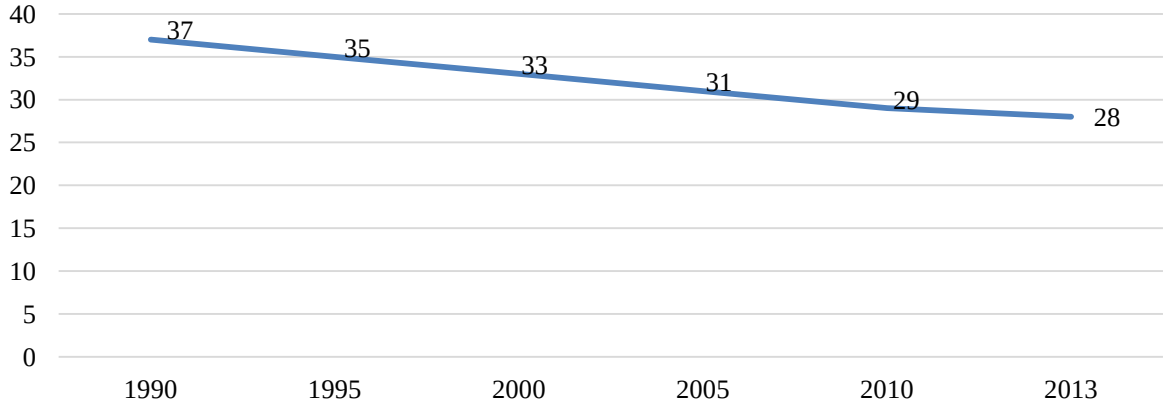
Note: "Although the World Bank standard is now US$1.25, the old reference of US$1 was the standard used in Nigeria at the time that the survey was conducted" National Bureau of Statistics (2012: 5) (note also that the 2011 figure was an estimate by the NBS).

Figure 9. Line Chart Nigeria's Employment to population Ratio (%)



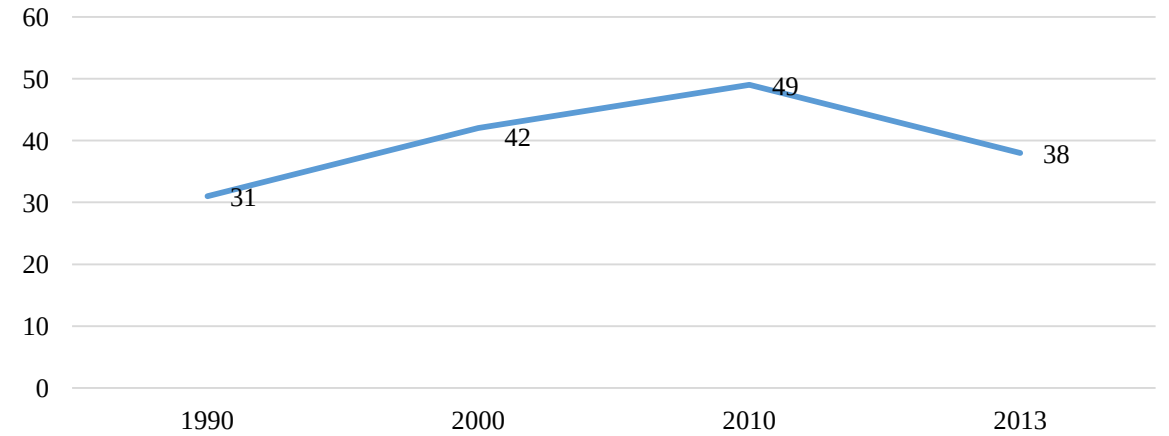
Source: author's line chart: with data from World Bank (2015c) Nigeria Millennium Development Goals, available from: http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=MDG-Table&Id=c658ae98&inf=n

Figure 10. Line Chart Nigeria's % Of Population with Access to Improved Sanitation Facilities



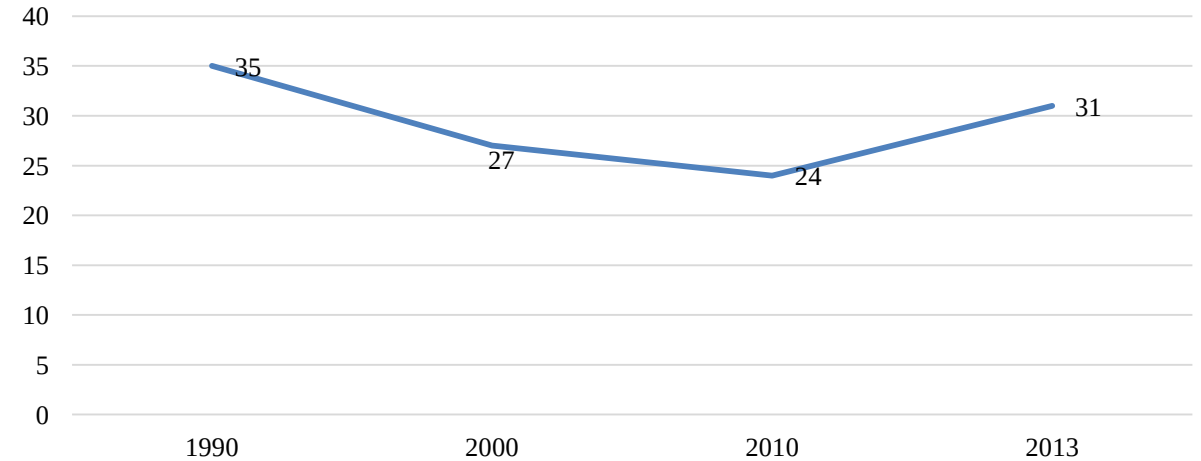
Source: author's line chart: with data from World Bank (2015c) Nigeria Millennium Development Goals, available from: http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=MDG-Table&Id=c658ae98&inf=n

Figure 11. Line Chart: Maternal Health, Nigeria's % Of Births Attended by Skilled Health Staff



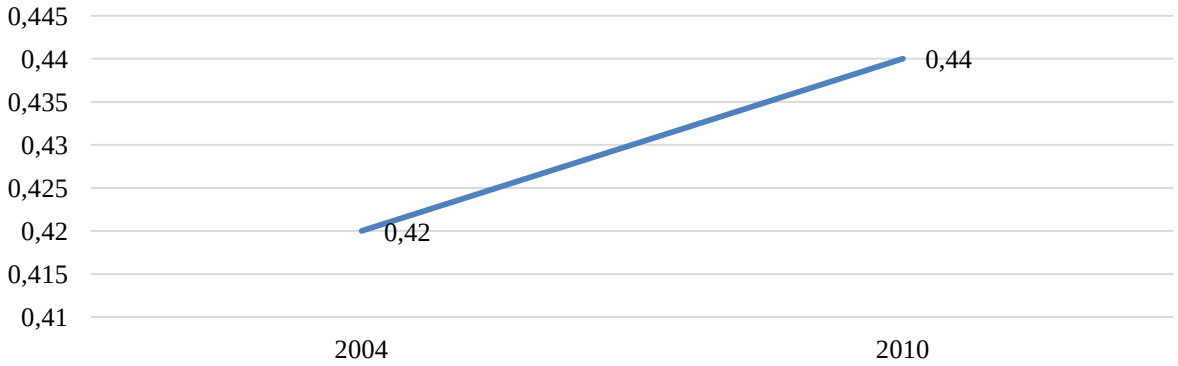
Source: author's line chart: with data from World Bank (2015c) Nigeria Millennium Development Goals, available from: http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=MDG-Table&Id=c658ae98&inf=n

Figure 12. Line Chart: Nigeria, Malnutrition prevalence (% of children under 5)



Source: author's line chart: with data from World Bank (2015c) Nigeria Millennium Development Goals, available from: http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=MDG-Table&Id=c658ae98&inf=n

Figure 13. Line Chart: Nigeria Income Inequality (GINI Index)



Source: author's line chart with data from: National Bureau of Statistics Nigeria (NBS) (2012) Nigerian poverty profile 2010 report, available at: http://reliefweb.int/sites/reliefweb.int/files/resources/b410c26c2921c18a6839baebc9b1428fa98fa36a.pdf

The poverty and income inequality implication of trade misinvoicing is eloquently substantiated by Jansky (2013). According to Jansky, trade misinvoicing schemes drains the government's revenue and reduces government's ability to provide public infrastructure – thus widening the income gap; it also reduces availability of private investment funds as money is shifted outside the country, this thus heightens unemployment –given low investment.

It is thus germane for policy makers to understand how trade misinvoicing has posed a canker to the fabric of Nigerian social and economic development. Within this period of an unprecedented peak in trade misinvoicing outflow from Nigeria, poverty levels increased around the year 2004 when a remarkable illicit financial outflow and associated trade misinvoicing outflow from Nigeria was recorded by the Global Financial Integrity (Global Financial Integrity, 2014). One of the key millennium development goals is the eradication of poverty, however, an obscured but yet perceivable harm meted out by trade misinvoicing to the Nigerian economy can be seen through the poverty ratios in Figure 8. Whilst the percentage of the Nigerian population living in poverty (less than $1 a day was 51.6% in 2004), the number of poor people climbed sharply to 61.2% in 2011 (National Bureau of Statistics, 2012). Yet the poverty numbers may appear somewhat neutralised, reason being that it does not really show how some parts of the country are poorer than others. For example, according to National Bureau of Statistics (2012), relative poverty was seen to be dominant in Northern part of the country with Sokoto State recording a shocking relative poverty level of 86.4% and other northern regions recording 77.7% and 76.3 percent. In addition, as at 2013, 31% percent of children under the age of five in Nigeria suffered from malnutrition, an increase from 24% in 2010 (Figure 12) – which is another indication of the extent of poverty.

Another important sustainable development indicator related to the poverty level is the employment level, Figure 9 shows that as at 2013, only about 52% of the Nigerian population were employed – showing that up 48% were unemployed. Another millennium development goal is to ensure environmental sustainability through citizens' access to improved sanitation facilities, but as at 2013, only about 28 percent of the Nigerian population had access to improved sanitation facilities. This implies that more than 70% of the population has little or no access to improved sanitation facilities. Figure 10 shows how the Nigerian population with access to improved sanitation has decreased over the years, corresponding to the years that the development fund has been drained outside Nigeria through illicit trade misinvoicing outflow. Furthermore, another sustainable development goal is to achieve reduction in maternal mortality through improved maternal health care. However, as at 2013, only about 38% of births were attended by skilled health staff (Figure 11). This suggests that more than 60% of expectant mothers had no access to improved maternal health benefits.

Furthermore, a growing refocus about the ideals of sustainability development has identified income inequality as an obstacle to sustainable development. Income inequality breeds economic and social instability and destroys the social structure and unity and hence becomes an impediment to the actualisation of environmental, socio-economic ethical behaviour to spur sustainable development. By draining a nation's financial resources off its shores through trade misinvoicing, a nation's income equality is likely to tilt in favour of the rich. Figure 13 characterises the Nigerian state of income inequality – shooting up from 42.0 (GINI Index) in 2004 during huge trade misinvoicing outflow from Nigeria and peaked at 44.0 in 2010 after the 2009 $17.9b trade misinvoicing outflow from Nigeria (Table 6).

"Gini index measures the extent to which the distribution of income or consumption expenditure among individuals or households within an economy deviates from a perfectly equal distribution…a Gini

index of 0 represents perfect equality, while an index of 100 implies perfect inequality" (World Bank, 2015d:1).

A GINI coefficient of 44.0 therefore implies that Nigeria's income and consumption distribution is heavily tilted with the majority having little to sustain daily living. The foregoing analysis epitomises the dwindling state of Nigeria's sustainable development and this is one of the development abrasions in an economy where heavy trade misinvoicing outflow is prevalent – showing that business and other entities engaged in trade misinvoicing have had a free illicit income shifting ride out of the Nigerian economy to tax havens (Tafirenyika, 2013; Global Financial Integrity, 2014).

On recognition of the sustainable development implication of trade misinvoicing, the United Nations Economic Commission for Africa (UNECA) has called for the inclusion of trade misinvoicing in the incoming post-2015 sustainable development goals, with a measurable achievable target for halting trade misinvoicing (Global Financial Integrity, 2015c). UNECA High Level Panel (HLP) on Illicit Financial Flows from Africa has called for, amongst others, that given the high percentage attributed to the trade misinvoicing component of illicit financial flows leaving Africa and other developing nations, a renewed effort with priority accorded to trade aspects of illicit flows from Africa and the rest of the world to curb trade misinvoicing (Global Financial Integrity, 2105a). Part of this effort should be the inclusion of trade misinvoicing in the new post-2015 sustainable development goals: "*the final SDGs should include a clear, concise, and measurable target to curb illicit financial flows from trade misinvoicing by 50% by 2030"* (Global Financial Integrity, 2015:3).

On their part, given that Nigeria is an unfortunate victim of illicit trade misinvoicing outflow with apparent developmental setbacks, Nigerian policy makers should urgently design a new strategy to stop trade misinvoicing outflow from Nigeria. Part of this strategy may include carving out a special wing from the current customs department and to empower this group with special education and skills on international trade misinvoicing. This special trade misinvoicing customs wing may thus be entrusted with the sole task of monitoring and halting trade misinvoicing outflow from Nigeria. An effort to halt trade misinvoicing may also be enhanced through tracking and registering of all businesses in Nigeria, including the beneficiaries. In this way, fund movements outside the country can easily be traced and queried.

7 Conclusion

This paper set out to examine how trade misinvoicing orchestrates external debt in Nigeria and the concomitant obstructive tendencies on Nigeria's sustainable economic development. The paper revealed that Nigeria is among the top ten developing countries who are victims of illicit trade misinvoicing outflows (Global Final Integrity, 2014:13) and render the country drained of its internal resources that should have been used for much needed developmental projects.

Findings show that trade misinvoicing outflow from Nigeria during the period 2003 – 2012 supersedes official development grants that were received in Nigeria. Findings also indicate that as trade misinvoicing outflow increased during the period 2003 -2012, Nigeria's external debt increased yearly. Aside from a graphical and tabular presentation, the researcher applied a T-test of difference in means to check for the degree difference between trade misinvoicing, external debt and official development assistance. Results from the statistical test showed that $P<0.05$ in each case – indicating that the mean difference in trade misinvoicing outflow is significantly greater than the mean differences in external debt and official development assistance. This finding attests to the huge internal financial resources that Nigeria lost during the period 2003 - 2012 through illicit trade misinvoicing outflow – and may continue to be lost if this conundrum is unchecked. The analysis further disclosed that trade misinvoicing outflow has hampered Nigeria's stride to sustainable economic development. There has been an increase in the level of unemployment with attendant increase in the poverty head count ratio – the percentage of Nigerians living under $1 per day. Similarly, the number of children suffering from malnutrition increased, a very small percentage of expectant mothers received child birth care from qualified health staff, a small percentage of the Nigerian population has had access to improved sanitation facilities and income inequality, as measured by the GINI index, widened more during the period 2003 – 2012.

Trade misinvoicing outflow has drained billions of Dollars of internal revenue from Nigeria, and this should be alarming to policy makers. As Nigeria transits into a new democratic leadership, which also coincides with global transition to a post-2015 sustainable development agenda, it is apposite and exigent to prioritise policies to curb trade misinvoicing outflows from Nigeria. This will provide Nigeria with enabling domestic financial resources to successfully implement health, education, domestic investment, and infrastructural development projects to reduce unemployment, poverty and to progress toward a desirable sustainable economic development trajectory.

Nigeria should increase its participation in the global movement for automatic exchange of tax information. This will reduce tax havens and secret cartel activities that incubate illicit trade misinvoicing outflows from Nigeria. This thus calls for a concerted effort from the Nigerian government to join other countries in a new measure of curbing anonymous

shell companies which is a recognised conduit for illicit financial flows. This can be done by creating a public registry of all Nigerian companies; *"such registries are considered the gold standard in curtailing the nefarious use of anonymous entities ….., such a move would go a long way toward curbing the terrible flow of dirty money"* (Baker, 2014:4). The paper also suggests the creation of a special wing of the Nigerian Customs. Educate and empower this wing with international trade misinvoicing techniques and tracking skills to enable this wing to deal specifically with trade misinvoicing outflows from Nigeria. Effective efforts to achieve a lasting sustainable development in Nigeria would require less reliance on external debt, but improved dependence on the country's internal resources through halting the outflow and retaining trade misinvoicing revenues in Nigeria. Mobilising such internal resources from trade misinvoicing will also require that the government remains committed in its efforts to stop public sector venality – a fault-line and a driver that facilitates the movement of illicit finance out of the country (UNECA, 2013).

References

1. Aizenman, J. 2008. On the hidden links between financial and trade opening. *Journal of International Money and Finance*, 27(3), 372-386.
2. Atkas, Z. and Aldan, A. 2013. Import surveillance and over-invoicing of imports in Turkey, Working paper No: 13/01, Central Bank of the Republic of Turkey 2013, available at: http://www.tcmb.gov.tr/wps/wcm/connect/9ffc0c2e-a807-44dd-ba48-b8c9a8208c0a/WP1301.pdf?MOD=AJPERES&CACHEID=9ffc0c2e-a807-44dd-ba48-b8c9a8208c0a [accessed 12 April 2015].
3. Baker, R. 2014. Want sustainable development? Mobilize domestic resources, curtail illicit capital flows, available at: https://www.devex.com/news/want-sustainable-development-mobilize-domestic-resources-curtail-illicit-capital-flows-83488 [accessed April 11, 2015].
4. Boyce, J. K. & Ndikumana, L. 2001. Is Africa a net creditor? New estimates of capital flight from severely indebted sub-Saharan African countries, 1970-96. *Journal of Development Studies*, *38*(2), 27-56.
5. Buehn, A. & Eichler, S. 2011. Trade Misinvoicing: The Dark Side of World Trade. The World Economy, 34(8), 1263-1287.
6. Buenos Aires Herald. 2014. Over-invoicing imports, available at: http://www.buenosairesherald.com/article/174059/gov%E2%80%99t-accuses-ge-of-overinvoicing-imports [accessed April 12, 2015].
7. Christian Aid. 2013. Invested interests, the UK's overseas territories' hidden role in developing countries, available at: http://www.christianaid.org.uk/Images/Invested-Interests-Christian-Aid-tax-report.pdf [accessed April 10, 2015].
8. Christensen, J. 2012. The hidden trillions: Secrecy, corruption, and the offshore interface. *Crime, Law and Social Change*, *57*(3), 325-343.
9. Diener, E.D., Tay, L. and Oishi, S. 2013. Rising income and the subjective well-being of nations, *Journal of Personal and Social Psychology*, 104(2), 267 – 276.
10. Financial Action Tax Force. 2006. Trade based money laundering, available at: http://www.fatf-gafi.org/media/fatf/documents/reports/Trade%20Based%20Money%20Laundering.pdf [accessed April 12, 2015].
11. Gelma, A. 2011. Why tables are really much better than graphs. *Journal Computational and Graphical Statistics*, 20(2), 3-7.
12. Gelma, A., Pasarica, C. and Dodhia, R. 2002. Let's practice what we preach: turning tables into graphs, *American Statistician*, 56, 121-130.
13. Global Financial Integrity. 2014 Illicit financial flows from developing countries http://www.gfintegrity.org/wp-content/uploads/2014/12/Illicit-Financial-Flows-from-Developing-Countries-2003-2012.pdf [accessed March 15, 2015].
14. Global Financial Integrity. 2015. Illicit financial flows, available at: http://www.gfintegrity.org/issue/illicit-financial-flows/ [accessed March 28, 2015].
15. Global Financial Integrity. 2015b. Trade misinvoicing p 3, available at: http://www.gfintegrity.org/issue/trade-misinvoicing/ [accessed April 10, 2015]
16. Global Financial Integrity. 2015c. AU/UN Panel Report Prioritizes Curbing Trade-Related Illicit Flows; Calls for SDGs to Follow Suit, available at: http://www.gfintegrity.org/press-release/au-un-high-level-panel-report-prioritizes-curbing-trade-related-illicit-flows-calls-sdgs-follow-suit/ [accessed April 3, 2015].
17. Global Financial Integrity. 2013. Illicit financial flows from developing countries: 2002-2011 http://iff.gfintegrity.org/iff2013/Illicit_Financial_Flows_from_Developing_Countries_2002-2011-HighRes.pdf [accessed March 15, 2015].
18. Global Financial Integrity. 2010. $854 Billion Removed from Africa by Illicit Financial Flows from 1970 to 2008 http://www.gfintegrity.org/press-release/854-billion-removed-africa-illicit-financial-flows-1970-2008/ [accessed March 15, 2015].
19. Global Financial Integrity. 2014. Illicit financial flows from developing countries 2003-2012, Illicit Financial Outflows from the Top 10 Developing Economies, 2003-2012, available at: http://www.gfintegrity.org/wp-content/uploads/2014/12/Illicit-Financial-Flows-from-Developing-Countries-2003-2012.pdf [accessed April 10, 2015].
20. Gray, R. & Bebbington, J. 1998. *Accounting and the soul of sustainability: hyper reality, transnational corporations and the United Nations*. University of Dundee. http://www.apira2013.org/past/apira1998/archives/pdfs/24.pdf [accessed March 15, 2015].
21. International Trade Alert. 2015. International price profiling system, available at: http://www.internationaltradealert.com/products.asp [accessed April 12, 2015].
22. Kastellec, J. P. & Leoni, E. L. 2007. Using graphs instead of tables in political science. *Perspectives on politics*, *5*(04), 755-771.
23. National Bureau of Statistics Nigeria (NBS). 2012. Nigerian poverty profile 2010 report, available at:http://reliefweb.int/sites/reliefweb.int/files/resource

s/b410c26c2921c18a6839baebc9b1428fa98fa36a.pdf [accessed April 20, 2015].
24. OECD. 2014. Curbing illicit financial flows for domestic resource mobilisation and sustainable development in the post-2015 era, side event report, available at: http://www.oecd.org/development/pcd/UNGA%202014%20Event%20Report.pdf [accessed March 15, 2015].
25. OECD. 2015. Beyond GDP, measuring progress, true wealth and the wellbeing of nations, available from: http://ec.europa.eu/environment/beyond_gdp/index_en.html [accessed May 5, 2015].
26. Sklair, L. 2002. *Capitalism and its alternatives*. Oxford: Oxford university press.
27. Social Progress Imperative. 2014. Economic Growth Does "Not Automatically" Lead To Social Advancement: New Social Progress Index, available from: http://www.socialprogressimperative.org/system/resources/W1siZiIsIjIwMTQvMDMvMzAvMDQvMTAvMDMvODM2L1NQSTIwMTRfR2xvYmFsUHJlc3Nlci5wZGYiXV0/SPI2014%20GlobalPresser.pdf [accessed May 3, 2015]
28. Stephens, B. 2002. Amorality of Profit: Transnational Corporations and Human Rights, The. *Berkeley J. Int'l L.*, *20*, 45.
29. Tafirenyika, M. 2013. Illcit financial flows from Africa: track it, stop it, get it, available at: http://www.un.org/africarenewal/magazine/december-2013/illicit-financial-flows-africa-track-it-stop-it-get-it [accessed May 2, 2015].
30. The Fiscal Times. 2011. The Ten Largest Global Business Corruption Cases, available from: http://www.thefiscaltimes.com/Articles/2011/12/13/The-Ten-Largest-Global-Business-Corruption-Cases [accessed May 3, 2015]
31. UNECA [United Nations Economic Commission for Africa]. 2013. The State of Governance in Africa: The Dimension of Illicit Financial Flows as a Governance Challenge, United Nations Economic and Social Council –Economic Commission for Africa, Third Meeting of the Committee on Governance and Popular Participation, Addis Ababa, Ethiopia 20 and 21 February 2013, available at: http://www.uneca.org/sites/default/files/uploaded-documents/CGPP/cgpp-3_illicit-financial-flow-english_final.pdf [Accessed March 15, 2015].
32. UNECA. 2013. The State of Governance in Africa: The Dimension of Illicit Financial Flows as a Governance Challenge http://www.uneca.org/sites/default/files/uploaded-documents/CGPP/cgpp-3_illicit-financial-flow-english_final.pdf [accessed March 15, 2015]
33. UN. 2013. The negative impact of the non-repatriation of funds of illicit origin on the enjoyment of human rights, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session22/A-HRC-22-42_en.pdf [accessed Feb 2, 2015].
34. World Bank. 2014. IBRD loans and IDA credits (DOD, current US$) available at: http://data.worldbank.org/indicator/DT.DOD.MWBG.CD [accessed Feb 2, 2015].
35. World Bank. 2015a. Net official development assistance and official aid received (current US$), http://data.worldbank.org/indicator/DT.ODA.ALLD.CD [accessed May 2, 2015].
36. World Bank. 2015b. Country and regional specific forecasts and data: GDP and Current Account Balance growth forecasts by country, region, and income level, available at: http://www.worldbank.org/en/publication/global-economic-prospects/data?region=SST [accessed May 2, 2015]
37. World Bank. 2015c. Nigeria Millennium Development Goals, available from: http://databank.worldbank.org/data/Views/Reports/ReportWidgetCustom.aspx?Report_Name=MDG-Table&Id=c658ae98&inf=n [accessed May 2, 2015]
38. World Bank. 2015d. GINI index (World Bank estimate), available from: http://data.worldbank.org/indicator/SI.POV.GINI [accessed May 2, 2015]
39. Wu, R. 2010. "Transfer Pricing: Current Problems and Solutions", CMC Senior Theses. Paper 87. http://scholarship.claremont.edu/cmc_theses/87 [accessed April 20, 2015]

IMPROVEMENT OF OPERATIONAL RISK MEASUREMENT UNDER THE SOLVENCY II FRAMEWORK

D. Stepchenko*, G. Pettere**, I.Voronova*

Abstract

Operational risk is one of the core risks of every insurance company in accordance to the solvency capital requirement under the Solvency II regime. The target of the research is to investigate the improvement possibilities of the operational risk measurement under Solvency II regime. The authors have prepared the algorithm of the operational risk measurement under Solvency II framework that helps improve the understanding of the operational risk capital requirements. Moreover, the authors have prepared the case study about a practical usage of the suggested algorithm through the example of one non-life insurance company. The authors use, in order to perform the research, such corresponding methods as theoretical and methodological analysis of scientific literature, analytical, statistical and mathematical methods.

Keywords: Modelling, Operational Risk, Skew t-copula, the Solvency II Framework

**Faculty of Engineering Economics and Management, Riga Technical University, Riga, Latvia*
Tel: +371-2214-6683
***Department of Engineering Mathematics, Riga Technical University, Riga, Latvia*
****Faculty of Engineering Economics and Management, Riga Technical University, Riga, Latvia*
Tel: +371-2633-0355

1 Introduction

Insurance in one of the most important areas in every country's economics therefore it requires more sophisticated and sensitive risk evaluation in order to ensure stability and solvency of an insurance company.

The Solvency II Framework has been under constant development for many ages due to the necessity of new approaches to ensure a more sensitive and sophisticated measurement, management and assessment of risk. In accordance with the Solvency II Directive's requirements, the insurance companies of the European Union should establish an effective risk assessment system with the aim to ensure policyholders' interests safety and the ability to prosper within the tough market environment.

The fact is that the Solvency II regime sets a lot of challenges to every insurance company since there is a need to seek for new approaches of risk measurement and their implementation in its processes and organizational structure.

The major problem lies in the fact that the Solvency II Directive's requirements are still under discussions and for that reason it is difficult to understand how to assess the risk.

The target of this research is to study the improvement possibilities of the operational risk measurement under the Solvency II regime.

The object of this paper is a measurement of operational risk. Therefore, the subject of this paper is the improvement of operational risk measurement using the skew *t*-copula.

In order to achieve the set objective, the authors use a theoretical and methodological analysis of scientific literature**,** as well as statistical and mathematical methods.

The main issue within the process of conducting the research was to interconnect the risk management with the risk measurement in an insurance company with the aim to improve risk assessment.

The article encompasses three main sections. The overview of the suggested improvement of operational risk measurement is presented in Section 2. In Section 3, the authors of the paper introduce the case study of enhancement of operational risk measurement in an insurance company through modelling. The final section summarizes the findings and conclusions of the research and assesses the improvement of operational risk measurement.

2 Improvement of risk measurement in insurance company

The Solvency II Directive is based on the three-pillar approach where each pillar fulfills its own function: quantitative requirements, qualitative and supervision requirements, disclosure requirements that mean prudential re-porting and public disclosure (FAQs, 2007).

Operational risk (OR) is the risk of a loss resulting from inadequate or failed internal processes,

people and systems, or from external events. This definition includes legal risk but excludes strategic and reputational risk (Embrechts and Hofert, 2011).

In the Solvency II framework and the Basel II regime, the basic principles and requirements for operational risk assessment in insurance and banking industries are described.

The usage, integration and implementation of the suggested principles as well as the requirements of operational risk assessment are under active discussions in the latest years. Many researchers (Embrechts and Hofert, 2014; Embrechts and Puccetti, 2008; Dutta and Perry, 2006; El-Gamal et al., 2007; Peters et al., 2013; Peters and Shevchenko, 2013; Frachot et al., 2001; Strelkov, 2008) are investigated those issues.

In the latest researches, in order to model operational risk losses depending on covariates to use an extension of the Peaks-over-Threshold method and the block maxima approach to a non-stationary setup that allows the dependence (on covariates) to be parametric, non-parametric, or semi-parametric and can also include interactions (Chavez-Demoulin et al., 2014).

Moreover, most heated discussions are going on in relation to the possibility that the capital, to cover the possible losses of the operational risk, can be directly proportional to the volume of gross profit in banking industry.

Thus, traditionally it is assumed that the amount of the capital, to cover the possible losses of the operational risk, is equal to the sum of capital charges for each type of the incurred unexpected event in insurance. However, the described approach requires an ideal dependence among the occurred events, which is unreasonable and unrealistic in business conditions of insurance industry.

The authors of the article suggest using copulas to model the capital volume to cover the operational risk. In fact, copulas allow to model multivariate probability distribution using one-dimensional parametric dependences. The fact is that copulas are used to describe the dependence between random variables. Actually, the copula's function enables the task of specifying the marginal distribution to be decoupled from the dependence structure of variables.

Consequently, copula's function allows us to exploit univariate techniques at the first step, and secondly, is directly linked to non-parametric dependence measures. This avoids the flaws of linear correlation that have, by now, become well known. (Cherubini et al., 2004)

Many authors have applied the difference copulas approaches to model the capital to cover the risks and other financial processes (Nelsen, 1999; Angela et al., 2009; Cherubini et al., 2004; Kollo and Pettere, 2010; Srelkov, 2009).

To model a capital to cover the operational risk the authors use skew *t*-copula. Skew *t*-copula is constructed from a multivariate skewed distribution, which has the covariance matrix when the number of degrees of freedom is more than 4 (Kollo and Pettere, 2009). Actually, this enables to model distributions with heavier tail area.

Since the operational risk encompasses a number of sub-risks, the authors suggest establishing the risk catalogue to investigate more deeply the nature of risks. Basically, the scope of risks that should be included in the analysis will depend on the purpose and context of the assessment (QIS5, 2010).

The authors of the paper suggest for the statistics for modelling the capital to cover the operational risk to use historical data from loss database. The fact is that loss database introduces all incurred operational risk events with details about the losses within a concrete period.

Loss databases, both internal and external, are important aspects of an operational risk program. An understanding of interconnectivity of different risks is a prerequisite to controlling problems and assessing practices. Firms should strive to understand the causes and related factors relevant to operational risk losses. Comprehensive qualitative information can help managers identify the commonalties among loss events. Seeing these patterns or common threads may allow managers to recognize red flags in their own controls before incidents occur. Quantitative tools further enhance a database by allowing it to be used for benchmarking (IAFE, 2011).

The authors of the paper have prepared the algorithm of measurement of the capital to cover the operational risk (see Figure 1).

The fact is that the authors of the paper have prepared the case study based on the algorithm presented in Figure 1 (see Section 3).

3 Case study: Assessment of operation risk

Due to the nature of operational risk that is less depended on macroeconomical cycles it can be modelled by skew *t*-copula and estimated tail dependence in each situation for modelling distributions with heavier tail area. The main idea of the case study is to approve the possibility of identification of VaR (the acronym standing for Value at Risk) for the operational risk portfolio using simulation technique.

Because of the correlation between different operational risk sub-risks, VaR of them (portfolio) has to be smaller than simply added corresponding VaR of each sub-risk.

The fact is that VaR is a quintile of a distribution and used as a (non-coherent) risk measure (CEA.2007).

The model created by the authors includes the following operational risk sub-risk:

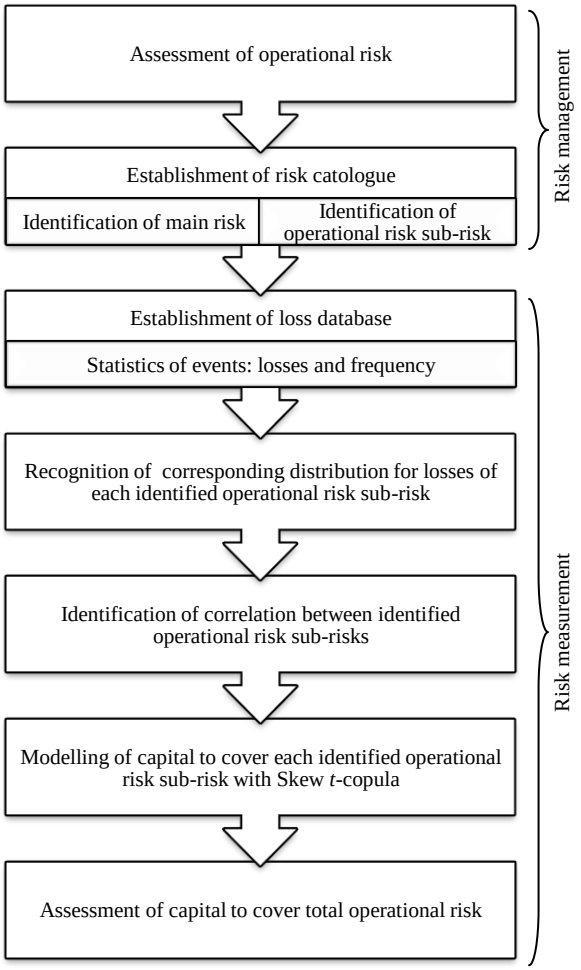
Legal risk (LR) means the possibility that lawsuits, adverse judgments from courts, or contracts that turn out to be unenforceable, disrupt or adversely

affect the operations or condition of an insurer. The result may lead to unplanned additional payments to policyholders or that contracts are settled on an unfavorable basis, e.g. unrecoverable reinsurance (CEA. 2007).

Organizational risk (OR) means possible losses due to unclear organizational structure (unclear processes, unclear responsibilities split between units etc.).

Informational risk (IR) means possible losses due to failures in the IT system.

Figure 1. The algorithm of measurement of the capital to cover the operational

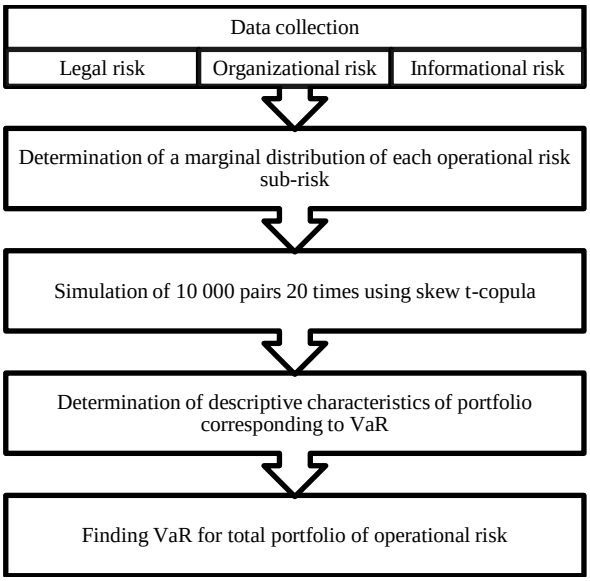


The simulation model performed during case study is based on three risks due to the reason to show the model advantages. In reality the proposed model is possible to use for any number of risks. The historical data was based on recorded data in relation to three operational risk sub-risk from the annual loss database.

Basically, the model is based on several main steps that are in details described in Figure 2.

However, the simulation of 10 000 pairs 20 times in the model is performed using the skew t-copula described in (Kollo and Pettere, 2010, 2011).

Figure 2. The description of model for determination of VaR for an operational risk portfolio



The correlation matrix is following:

$$R = \begin{pmatrix} 1 & -0.143 & 0.357 \\ -0.143 & 1 & -0.118 \\ 0.357 & -0.118 & 1 \end{pmatrix}$$

The legal risk and informational risk (first and third) are positively correlated but the other are negative. Descriptive statistics of the marginal distributions of the mentioned risks are presented in Table 1.

The fact is that, before fitting to marginal distributions, the data was standardized and then the marginal distributions were approximated by Exponential and Gamma distributions.

Table 1. Descriptive Statistics of Data

Risks	LR	OR	IR
Size	12	12	12
Mean	7 564	45 618	5 425
Median	3 700	1 610	960
Standard deviation	11 151	143 207	9 342
Skewness	3	3	2
Kurtosis	9	12	5

The authors of the paper have identified that the legal risk should be obtained by the Exponential distribution, but for the organizational and informational risk, the Gamma distribution should be suitable.

The appropriateness of distributions to each sub-risk was measured by the Kolmogorov test (the 5% critical value equals 0,391).

The testing results are shown in Table 2.

Table 2. Results of Marginal Distributions Tests

Risks	Used distribution	Parameters	
LR	Exponential	λ	1.474
		Test value	0.164
OR	Gamma	α	0.101
		β	3.139
		Test value	0.169
IR	Gamma	α	0.227
		β	2.098
		Test value	0.0957

Based on Table 2, the authors can draw conclusion that all univariate marginal distributions are appropriate to the obtained model distributions.

The obtained marginal distributions were joined into a three-dimensional distribution by the skew *t*-copula.

The parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ are estimated from the first two moments (Kollo and Pettere, 2010).

Let $\bar{\mathbf{X}}$ and $\mathbf{S}_{\mathbf{X}}$ denote the sample mean and the sample covariance matrix, respectively. Then the estimates are

$$\hat{\boldsymbol{\Sigma}} = \frac{\nu - 2}{\nu}(\mathbf{S}_{\mathbf{X}} + \bar{\mathbf{X}}\bar{\mathbf{X}}^T)$$

$$\hat{\boldsymbol{\alpha}} = \frac{b(\nu)\cdot\boldsymbol{\beta}}{\sqrt{b^2(\nu) - \bar{\mathbf{X}}^T\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{X}}}}$$

where

$$\boldsymbol{\beta} = \frac{1}{b(\nu)}\hat{\mathbf{W}}\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{X}}$$

With $\hat{\mathbf{W}} = (\delta_{ij}\sqrt{\hat{\sigma}_{ij}})$, $i, j = 1,..., p$, where δ_{ij} is the Kronecker delta and

$$b(\nu) = \left[\frac{\nu}{\pi}\right]^{\frac{1}{2}} \cdot \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$$

The number of degrees of freedom ν was taken as 4 (four) in order to use the multivariate *t*-distribution with maximally heavy tail area. The $\hat{\boldsymbol{\Sigma}}$ matrix is following:

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.730 & 0.037 & 0.340 \\ 0.037 & 0.551 & 0.017 \\ 0.340 & 0.017 & 0.614 \end{pmatrix}$$

However, the calculated values of alfa are following:

$$\boldsymbol{\alpha}^T = (1.551 \;\; 0.946 \;\; 0.681)$$

In the experiment of simulation triples from the joint 3-variate skew *t*-copula were modelled. The number of replications was 20. The results of simulation are collected in Table 3. On the first line `Real values´ we have the 99.5% *VaR* for each sub-risk using inverse marginal distributions and sum of *VaR* (portfolio) in the current year.

On the next lines, characteristics of 99.5% *VaR* for each sub-risk and portfolio obtained from modelled simulations.

Table 3. 99.5% *VaR* Obtained Using Simulation and its Characteristics

Risks	LR	OR	IR	Sum of *VaR*	Portfolio
99.5% VaR from distributions	40 078	947 292	55 567	1 042 937	
Mean of *99.5% VaR*	40 091	909 123	56 556	1 005 769	935 922
Median	40 034	91 1132	56 821	1 008 493	935 630
Standard deviation	1 005	41 170	2 888	42 721	4 4248
Skewness	0.232	-0.008	-0.399	-0.035	0.178
Coefficient of variation (%)	2.5	4.5	5.1	4.2	4.7

Based on Table 3, it is possible to conclude that the portfolio *VaR* obtained in simulation is smaller, and it means that the capital to cover these risks is less by 10.3%.

In order to evaluate the dependence between risks the authors have used tail dependence coefficient (Bortot, 2012).

Let assume that (X_1, X_2) is two dimensional vector with one dimensional marginal distributions $F_1(x)$ and $F_2(x)$. Then the upper tail coefficient is

$$\lambda_U = \lim_{u\to 1}\lambda_U(u)$$

where $\lambda_U(u) = P(F_1(x) > u / F_2(x) > u)$.

Similarly is defined lower tail coefficient

$$\lambda_L = \lim_{u \to 1} \lambda_L(u)$$

where $\lambda_L(u) = P(F_1(x) < u / F_2(x) < u)$.

$\lambda_U = \lambda_L = \lambda$ for symmetrical elliptical distribution, but for normal distributions λ equals zero. For two dimensional *t*-distribution with ν degrees of freedom

$$\lambda = 2T_{1,\nu}\left(-\sqrt{\frac{(\nu+1)\cdot(\rho-1)}{(\rho+1)}}\right)$$

where

$T_{1,\nu}(.)$ - the distribution function of standard *t* - distribution with ν degrees of freedom;

ρ – coefficient of correlation.

It is approved in (Bortot, 2012) that it is sufficient to study the upper tail dependence due to the lower tail dependence coefficient that is determined by the upper one. Let us denote by

$$\alpha_1^* = \frac{\alpha_1 + \alpha_2 \cdot \rho}{\sqrt{1+\alpha_2^2 \cdot (1-\rho^2)}}.$$

and

$$\alpha_2^* = \frac{\alpha_2 + \alpha_1 \cdot \rho}{\sqrt{1+\alpha_1^2 \cdot (1-\rho^2)}}.$$

Assume that $\alpha_1^* \le \alpha_2^*$. Then

$$\begin{aligned}\lambda_U &= \lim_{u\to 1}\frac{P(F_1(x)>u, F_2(x)>u)}{P(F_2(x)>u)} = \lim_{x\to\infty}\frac{P(F_1(x)>F_2(x), X_2>x)}{P(X_2>x)} \\ &\ge \lim_{x\to\infty}\frac{P(X_1>x, X_2>x)}{P(X_2>x)} \\ &= \lim_{x\to\infty}\frac{2\cdot P(Y_1>x, Y_2>x)\cdot T_{1,\nu+2}\left((\alpha_1+\alpha_2)\cdot\sqrt{\frac{(\nu+2)\cdot(\rho+1)}{2}}\right)}{2\cdot T_{1,\nu+1}(\alpha_2^*\cdot\sqrt{\nu+1})\cdot(1-T_{1,\nu}(x))} \\ &= \lambda\cdot\frac{T_{1,\nu+2}\left((\alpha_1+\alpha_2)\cdot\sqrt{\frac{(\nu+2)\cdot(\rho+1)}{2}}\right)}{T_{1,\nu+1}(\alpha_2^*\cdot\sqrt{\nu+1})}\end{aligned}$$

In the case of $\alpha_1 = \alpha_2 = \alpha$ tail dependence coefficient can be calculated using formula:

$$\lambda_U = \lambda\cdot\frac{T_{1,\nu+2}\left(2\cdot\alpha\cdot\sqrt{\frac{(\nu+2)\cdot(\rho+1)}{2}}\right)}{T_{1,\nu+1}(\alpha^*\cdot\sqrt{\nu+1})},$$

where

$$\alpha^* = \frac{\alpha\cdot(1+\rho)}{\sqrt{1+\alpha^2\cdot(1-\rho^2)}}.$$

The fact is that the difference of tail dependencies between *t*-distribution and skew *t*-distribution is determined by the ratio of univariate distribution functions of the *t*-distribution. It is shown in (Kollo, Pettere, Valge, 2015).that for the equal values of α the difference in tail dependence is not large.

The tail dependence coefficient calculations for given risks is presented in Table 4.

Table 4. Results of Tail Dependence Coefficient for the Risk

Risks	Legal risk – Organizational risk	Organizational risk – Informational risk	Legal risk – Informational risk
λ	0.030612	0.032737	0.099198
α_1^*	1.033629	0.716906	1.513446
α_2^*	0.395529	0.414779	0.700886
$T_{1,\nu+2}$	0.996458	0.980886	0.997954
$T_{1,\nu+1}$	0.791520	0.801877	0.911078
λ_U	**0,038537**	**0.040046**	**0.108657**

The measurement of the operational risk based on copulas allow modelling multivariate probability distribution using one-dimensional parametric dependencies.

The measurement of the operational risk is based on the skew *t*-copula since it allows modelling distributions with heavier tail area and correlation between marginal distributions. However, there are discovered several valuable advantages of skew *t*-copula usage in operational risk measurement:

skew *t*-copula has a very simple and clear simulation rules;

using copula is possible to simulate portfolio of risks keeping correlation between them;

calculated necessary capital for portfolio is less than sum of capitals needed for each risk;

by choosing degrees of freedom is possible to find appropriate skewness of copula for simulation;

another advantage of simulation is the possibility to calculate average measure of necessary characteristic;

further tail dependence can be evaluated between risks.

During the case study, it has been proved that because of the correlation among different sub-risks of the operational risk, their *VaR* (portfolio) is smaller than a simply added corresponding *VaR* of each sub-risk that allows keeping optimal volume of capital to cover the possible losses due to occurrence of the operational risk. Because *VaR* is not coherent risk measure the *VaR* for simulated portfolio will always be less than sum of *VaR* of different risks. Thus, the proposed method would not allow over-reserving and putting gap capital to other needs of an insurance company.

4 Conclusion

The dynamic nature of risk under changing insurance market conditions sets a lot of challenges to every insurance company.

Moreover, the new Solvency II Directive's requirements, which will soon come in force, set a lot of challenges to every insurance company in the countries of the European Union in relation to the establishment of more sensitive and sophisticated risk coverage in order to ensure solvency to ensure the safety of the policyholders.

The fact is that the new regime requirements might create additional problems for an insurer.

The authors of the paper have interconnected the risk management with the risk measurement in an insurance company with the target to improve the operational risk assessment.

Basically, the authors suggest the algorithm of the operational risk evaluation to measure the capital to cover it. The measurement of the operational risk is based on copulas since they allow to model multivariate probability distribution using one-dimensional parametric dependencies.

Furthermore, the authors have prepared the case study in accordance with the suggested algorithm. The main idea of the case study is to approve that because of the correlation between different operational risk sub-risks, *VaR* of them (portfolio) has to be smaller than a simply added corresponding *VaR* of each sub-risk.

The suggested approach of the capital measurement to cover the operational risk will enable every insurance company to control and properly assess the capital required for the operational risk in line with the Solvency II Directive requirements and establish a more sophisticated and sensitive risk assessment in future.

In future, the authors plan to continue the present research on an insurance company's risk assessment.

References

1. Angela, C., Bisignani, R., Masala, G., Micocci, M. (2009). "Advanced operational risk modelling in banks and insurance companies". Investment Management and Financial Innovations. Vol. 6, Issue 3, pp.73-82.
2. Bortot P., Tail dependence in bivariate skew-Normal and skew-t distributions. 2012. / Internet. - ttp://www2.stat.unibo.it/bort 15 pp.
3. Cherubini, U., Luciano, E., Vecchiato, W. (2004). Copula methods in finance, Willey, New-York.
4. Chavez-Demoulin, V., Embrechts, P., Hofert, M. (2014). "An extreme value approach for modelling operational risk losses depending on covariaties". Journal of Risk and Insurance. Retrieved from http://www.math.ethz.ch/~embrecht/papers.html
5. Comité Européen des Assurances (CEA) and the Groupe Consultatif Actuariel Européen (Groupe Consultatif). (2007). Solvency II Glossary. Retrieved from http://ec.europa.eu/internal_market/insurance/docs/solven cy/impactassess/annex-c08d_en.pdf
6. Dutta, K., Perry, J. (2006). "A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital". Working paper. Federal Reserve Bank of Boston. Retrieved from http://www.bostonfed.org/economic/wp/wp2006/wp0613.pdf
7. El-Gamal, M., Inanoglu, H.. Stengel, M. (2007). "Multivariate estimation for operational risk with judicious use of extreme value theory". Journal of Operational Risk. Vol. 2(1), pp. 21-54.
8. Embrechts, P. & Hofert, M. (2014). "Statistics and Quantitative Risk Management for Banking and Insurance". Annual Review of Statistics and its Applications. Vol.1, pp. 493-514.
9. Embrechts, P. & Hofert, M. (2011). " Practices and issues in operational risk modelling under Basel II". *Lithuanian Mathematical journal – LITH MATH J.* Vol.52, N.2, pp. 180–193. DOI: 10.1007/s10986-011-9118-4
10. Embrechts, P. & Puccetti, G. (2008). "Aggregating risk across matrix structured loss data: the case of operational risk". Journal of Operational Risk, Vol. 3(2), pp. 29-44.
11. European commission. "'SOLVENCY II': Frequently Asked Questions (FAQs)". (2007). Retrieved from http://europa.eu/rapid/press-release_MEMO-07-286_en.htm?locale=en
12. European Insurance and Occupational Pensions Authority (2010). Technical specification for Quantitative impact study 5. QIS5. Retrieved from http://ec.europa.eu/internal_market/insurance/docs/solvency/qis5/201007/technical_specifications_en.pdf
13. Frachot, A., Georges, P., Roncalli, T. (2001). Loss Distribution Approach for Operational Risk, Working papers. Retrieved from http://thierry-roncalli.com/download/lda.pdf
14. International Association of Financial Engineers (IAFE). (2011). Report of the Operational Risk Committee: Evaluating Operational Risk Controls.
15. Kollo,T. & Pettere, G.(2010). "Parameter Estimation and Application of the Multivariate Skew t-Copula". Copula Theory and Its Applications. Proceedings of the Workshop Held in Warsaw, 25-26 September

2009. Springer, pp. 289-298. (MathSciNet, Zentralblatt MATH).
16. Nelsen, R. (1999). An introduction to copulas. Springer-Verlag, New-York.
17. Peters, G.W., Targino, R., Shevchenko, P. (2013). "Understanding Operational Risk Capital Approximations: First and Second Orders". *Journal of Governance and Regulation*, 2(3), pp.58-78.
18. Peters, G.W., Shevchenko, P. (2013). "Loss Distributional Approach of Operational Risk Capital Modelling under Basel II: Combining Different Data Sources for Risk Estimation". *Journal of Governance and Regulation Vol. 2(3), pp. 37-57.*
19. Pettere, G. & Kollo, T. (2011). "Risk Modeling for Future Cash Flow Using Skew t-Copula". Communications in Statistics – Theory and methods. pp. 2919-2925
20. Kollo T., Pettere G., Valge M., Tail Dependence of Skew t-Copulas. - Communications in Statistics, 2015, accepted.
21. Стрелков, С. (2008). Страхование операционных рисков: применение теории разорения для расчета величины достаточного капитала, *XV Международная конференция «Ломоносов-2008», Экономика.* МГУ, Retrieved from http://lomonosov-msu.ru/archive/Lomonosov_2008 /prog ram_2008.pdf
22. Стрелков, С. (2009). Страхование операционных рисков: расчет величины достаточного капитала в моделях Stop-Loss. *XVI Международная конференция «Ломоносов-2008», Экономика.* МГУ. pp.27-29. Retrieved from http://lomonosov-msu.ru/archive/Lomonosov _2009/29_4.pdf

OPERATIONAL RISK IN BANK GOVERNANCE AND CONTROL: HOW TO SAVE CAPITAL REQUIREMENT THROUGH A RISK TRANSFER STRATEGY. EVIDENCES FROM A SIMULATED CASE STUDY

E. Scannella*, G. Blandi*

Abstract

Operational risk management in banking has assumed such importance during the last decade. It has become increasingly important to measure, manage, and assess the impact of operational risk in the economics of banking. The purpose of this paper is to demonstrate how an effective operational risk management provides mitigating effects on capital-at-risk in banking. The paper provides evidences that an implementation of an operational risk transfer strategy reduces bank capital requirement. The paper adopts the loss distribution approach, the Monte Carlo simulation, and copula methodologies to estimate the regulatory capital and simulate an operational risk transfer strategy in banking***.

Keywords: Operational Risk, Risk Transfer, Banking, Basel Accord, Risk Management, Financial Regulation

JEL Classification: G01, G2, G18, G21, G24, G28, G32

**Department of Economics, Business, and Statistics, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy*
Tel: +39 091 23895305
Fax: +39 1782217366
***Department of Economics, Business, and Statistics, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy*
***Although the paper has been written jointly by the two Authors it is possible to identify the contribution of each one as follows. Abstract, and sections 1, 2, 3, 6 have been written by Enzo Scannella. Sections 4 and 5 have been written jointly by Enzo Scannella and Giuseppe Blandi. The data were analysed jointly by the two Authors. All the figures and tables were prepared jointly by the two Authors.*

1 Introduction

This paper aims to demonstrate how an effective operational risk management provides mitigating effects on capital-at-risk in banking. The paper provides evidences that an implementation of an operational risk transfer strategy saves capital requirement and reduces the cost of capital in banking. To estimate the regulatory capital the paper adopts an advanced measurement approach, and particularly the loss distribution approach. Such approach is based on a bottom-up methodology. Then, the analysis is conducted on a simulated operational losses database. The estimation of the loss distribution has been carried out using Monte Carlo simulation and copula methodologies.

The analysis is carried out in two parallel steps. In the first one, the operational risk capital requirement is estimated. In the second one, an operational risk transfer policy is implemented through the insurance market. Such policy provides a mitigating impact on the regulatory capital.

The structure of this paper is as follows. Section II introduces operational risk in banking. It aims to frame the specific nature of operational risk. Section III provides a regulatory perspective of the operational risk with reference to the first, second, and third pillar of the New Basel Capital Accord. Section IV analyses the loss modelling process that is based on a separate estimation of the frequency distribution and severity distribution of a single operational event. Section V provides a simulated operational losses database that supports the operational risk transfer strategy in banking. Section VI concludes.

2 An introduction to operational risk in banking

The operational risk is one of the most important risk in the economics of banking. It is defined for the first time by the Basel Committee on Banking Supervision (2001) as "the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events". This definition has been incorporated into the New Bank Capital Accord (Basel Committee on Banking Supervision, 2006). The Basel Committee's definition incorporate the legal risk but excludes the reputational and strategic

risks.

Under the 1988 Capital Accord, there were no capital buffers for such kind of risk. Only with the New Capital Accord in 2006 the Basel Committee recognizes the importance of operational risks, in addition to credit and market risks.

Operational risk management in banking has assumed such importance during the last decades. It has become increasingly important to measure, manage, and assess the impact of operational risk in the economics of banking. Operational risk events may have considerable economic consequences in banking that could compromise the business continuity. As stated by the Basel Committee on Banking Supervision (2001, p. 1) "developing banking practices such as securitization, outsourcing, specialized processing operations and reliance on rapidly evolving technology and complex financial products and strategies suggest that these other risks are increasingly important factors to be reflected in credible capital assessments by both supervisors and banks".

The definition is based on the underlying causes of operational risk. Briefly, the drivers of operational risk are: internal processes, people, information systems, and external events. The operational risk comes from very different causal factors (event risk) and it is inextricably linked to bank activities. From an organization point of view, this kind of risk is pervasive, transversal and similar to a "pure risk". A greater risk is not associated with a higher expected return. The operational risk is not taken in return for an expected reward, like financial risks.

Operational risk is a kind of risk that affects all financial institutions (Santomero and Babbel, 2012). Operational risk is a normal part of banking. There is, however, a trade-off problem in defining an appropriate balance between the benefit of eliminating the risk and the cost of the risk reduction/mitigation (Bessis, 2009). The principles for the management and supervision of operational risk issued by the Basel Committee on Banking Supervision (2011a) recognize that it is essential that banks have a comprehensive risk management process in place that effectively identifies, measures, monitors and controls operational risk exposures, and that is subject to appropriate board and senior management oversight. Sound risk management practices are essential to the prudent operation of banks and the stability of the financial system. A sound risk management process may be divided into four steps. The first one is the identification and understanding of operational risk. The second step is the analysis and the identification of the drivers and principal components of the operational risk in banking. The next step is the measuring of the operational risk, using different models and approaches that are available for different kinds of banks. A bank has to balance between the cost of using a model and the benefits in terms of quality and reliability of risk measures. The final step is the management of operational risk, in order to reduce/mitigate or eliminate the impact of the operational rate risk in the economics of banking.

Briefly, the introduction of Basel II is important not only because it imposes some standards methodologies for assessing the operational risk capital requirement in banking, but also because it predicts radical changes in the management structures and processes in banking (Birindelli and Ferretti, 2006, 2009; Resti and Sironi, 2007, 2008; Scannella, 2005; Sironi, 2003). A proactive operational risk management, a strong involvement of the top management, a constant auditing activity by the bank Internal Audit, a recurrent review of the operational risk management processes, and well-defined reporting systems and responsibility frameworks at business unit levels are all principles for an effective operational risk management implementation in banking.

3 Operational risk in banking: a regulatory perspective

The New Basel Accord (Basel Committee on Banking Supervision 2006) introduces a capital charge for operational risk in banking. The Basel Committee on Banking Supervision established a minimum regulatory capital charge for operational risk under Pillar 1.

The New Basel Accord presents three methods for calculating operational risk capital charges in a continuum of increasing sophistication and risk sensitivity. The availability of different methodologies aims to ensure correspondence between the complexity of the approaches, and the improvements of risk management practices in banking (Gabbi et al., 2005; Hull, 2012).

The basic approach to measure operational risk is the Basic Indicator Approach. It uses a single indicator as a proxy for the overall operational risk exposure. The bank capital requirement is determined applying a 15% coefficient to the average of the last three years' positive annual gross income. It is an extremely easy approach to implement across banks. A bank's gross income is the only component that is taken into consideration to evaluate the operational risk capital charge.

The second approach is the Standardised Approach. The main difference from the first approach is that a bank's activities are divided into a number of standardised business units and business lines. For each of them it is associated a beta coefficient, which is then multiplied by the last three years' average gross income. The resulting operational risk capital requirement is then obtained as a sum of each business line. Differences in the beta coefficients are linked to the different impact of operational losses on the income capacity of each business lines.

Within each business line, the capital charge is calculated by multiplying a bank's financial indicator

by a "beta factor". It represents a rough proxy for the relationship between the industry's operational risk loss experience for a given business line and the broad financial indicator representing the banks' activity in that business line. The gross income serves as a scale of operational risk exposure within each business line.

The differentiation among business lines represents a step forward in comparison to the basic approach. Nevertheless, the Standardised Approach is also affected by many simplifying assumptions: the existence of a perfect correlation among different loss events, mitigation policies are neglected, extreme events are not caught, and the gross income of each business line is a rough proxy of the bank's risk exposure. In addition, it does not seem to foster the development of appropriate techniques and strategies to face effectively up operational risks in banking.

The third approach is the Advanced Measurement Approach. This approach, in comparison to the previous two, is much more complex and requires qualitative and quantitative standards, in terms of organizational requirements, effective internal control mechanisms and operational risk management techniques. The Advanced Measurement Approach is based on the estimation of a loss frequency and loss severity distribution. The estimation is supported by internal and external historical data. This approach aims to quantify the operational risk exposure, without using any kind of proxy, and differentiate it by business lines. It recognises that the operational risk is the result of two factors: the probability that an event will occur and the consequences of the adverse event. Banks are allowed to implement risk mitigation strategies, use risk transfer mechanisms, and hedge risk exposure with insurance policies (Basel Committee on Banking Supervision, 2003). The Advanced Measurement Approach requires Value at Risk methodologies (Operational VaR) to evaluate the operational unexpected loss, using a 99,9% confidence level and a 1-year time horizon (Basel Committee on Banking Supervision, 2006, 2011b).

In order to implement an Advanced Measurement Approach banks need to categorize operational risks and business lines, according to the Basel Committee-specified event types and business lines. The categorization of operational risk is as follows: internal fraud; external fraud; employment practices and workplace safety; clients, products and business practices; damage to physical assets; business disruption and system failures; execution, delivery and process management. The categorization of business lines is as follows: corporate finance; trading and sales; retail banking; commercial banking; payment and settlement; agency services; asset management; retail brokerage. Banks need to estimate their exposure to each combination of type of risk and business line. Ideally this will lead to 7×8=56 VaR measures that can be combined into an overall VaR measure.

The Advanced Measurement Approach provides incentives for banks to develop measurement methodologies and techniques to internally estimate operational risk and calculate regulatory capital requirements. The Advanced Measurement Approach is the most risk sensitive of the approaches currently being developed for regulatory capital purposes. As market risk capital requirements, the operational risk capital requirements are based on internal models that are developed by banks. These models are subject to qualitative and quantitative standards. They use internal and external loss data (industry loss data).

Nevertheless the above mentioned advantages, the Advanced Measurement Approach is affected by several concerns and criticisms, such as the difficulties to measure the operational risk, the complexity of the calculations, and the non-normal distribution of loss frequency and loss severity (Birindelli and Ferretti, 2006, 2009; Resti and Sironi, 2007, 2008; Tutino, Birindelli and Ferretti, 2011, 2012).

A key issue in the development and implementation of regulatory capital requirements and internal approaches to measure the operational risk is the collection and analysis of loss data, as well as the definition of industry standards to share loss data across banks. Banks need to develop advanced information systems to support an internal measurement approach for operational risk management (Aprile, 2007; Cosma, 2008; Gabbi et al., 2005).

Banks are encouraged to develop sophisticated techniques and practices to manage and monitor their operational risks. The financial regulation issued by the Basel Committee on Banking Supervision aims not only to ensure that banks have adequate capital to support risks (Pillar 1), but also to ensure that banks improve internal control processes, methodologies, and practices to increase the effectiveness of the operational risk management (Pillar 2). Banks have to identify and strengthen policies and strategies that support the assessing, monitoring and controlling/mitigating the operational risk, and establish adequate internal systems for measuring, monitoring, and reporting operational risk exposures. Pillar 2 recognizes that the risk faced by a bank depends on qualitative aspects, such as: organizational structure, internal control systems, and risk management practices.

Supervisors review and evaluate banks' internal capital adequacy assessment and strategies, as well as their ability to monitor and ensure their compliance with regulatory capital requirements. Briefly, capital ratios are not more important than the adequacy and effectiveness of operational risk management practices in banking. The qualitative analysis of the operational risk in banking is put at the center of the Internal capital adequacy assessment process (ICAAP) and the Supervisory review and evaluation process (SREP). There is a strong interlinking

between ICAAP and SREP in banking. It recognizes the relationship that exists between the amount of regulatory capital that a bank has to hold against its operational risk and the strength, soundness, and effectiveness of a bank's risk management and internal control processes. In that view, the ICAAP and SREP are complementary in banking. They are parts of a wider supervisory review process covered by Pillar 2.

To complement the capital requirements and the supervisory review process, the Basel Committee developed a set of risk disclosure requirements (Pillar 3) that aims to remove obstacles that prevent market discipline, and inform the market about a bank's risk exposure. Pillar 3 provides a disclosure framework based on qualitative and quantitative disclosure requirements. Banks are required to disclose: scope and application of Basel regulation; nature of capital held; regulatory capital requirements; risk management objectives, policies, processes and structures; nature of banks' risk exposures. The market discipline of Pillar 3 addresses the issues of transparency in banking.

4 Modelling operational losses in banking

In this section the paper aims to analyse an advanced measurement approach, and particularly the "loss distribution approach", to estimate the regulatory capital requirement in banking. Such approach is based on a bottom-up methodology in which operational loss data coming from internal databases, external consortiums, public data, and scenario analysis are used to develop an assessment activity at every business process in order to identify and quantify all types of operational risks (Alexander, 2003; Frachot et al., 2001). The loss distribution approach is characterized by the following steps:

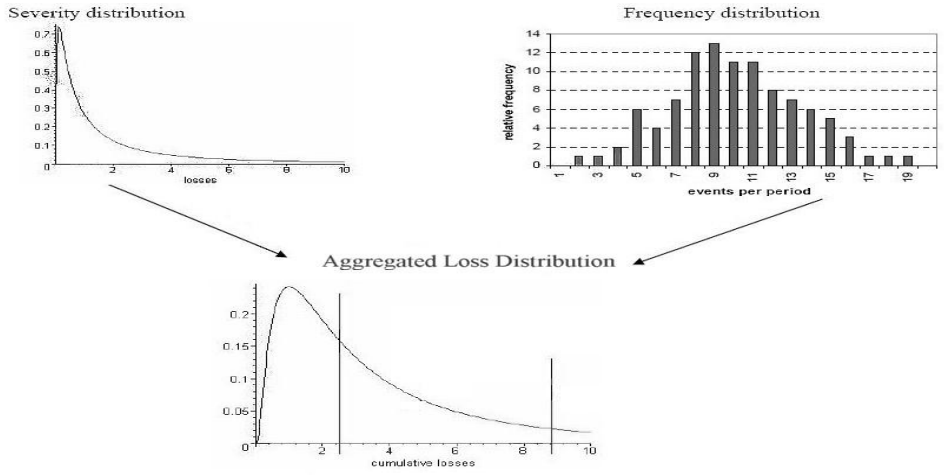
- risk class definition: operational risk data are classified in homogeneous categories, in such a way to satisfy the independence and identical distribution hypotheses. The number of classes determines greater or less granularity of the model. The financial regulation requires to test such hypothesis. Collecting data for event type and business line could be considered as a minimal risk class.

- estimation of the severity of operational loss: it identifies a monetary loss caused by an operational event. In order to estimate such severity it is necessary to select a list of possible distribution functions, find out the parameters that best match the observed data to the distribution, and test the distribution functions in order to select the best model.

- estimation of the frequency of operational loss: it is necessary to determine the distribution function that represents the number of observed operational events. The probability distribution should fit the empirical data.

- aggregation of severity and frequency distributions to obtain the aggregate loss distribution. For each risk class, it is necessary to compound severity and frequency into one aggregated loss distribution (Figure 1). It allow to forecast operational losses with a certain degree of confidence.

Figure 1. Aggregated loss distribution



- aggregation of the loss distribution of each risk class to determine the overall annual loss distribution. Adopting a conservative approach (it is based on the assumption of a perfect linear correlation among different risk classes), the operational risk capital requirement is the sum of the Capital-at-Risk for each risk class. However, the Basel Accord allows to use other aggregating techniques that better take into account the correlation among different risk classes.

- estimation of capital-at-risk: from a regulatory perspective the capital-at-risk is determined as the Value-at-Risk of the overall annual loss distribution with a 99,9% confidence level.

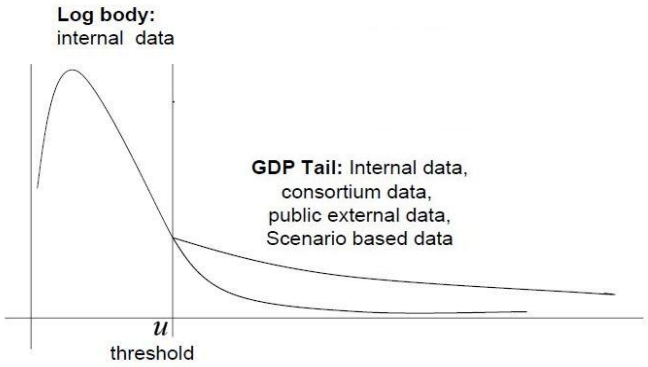
Let us analyse some details of the above

mentioned steps in the following sub-sections.

4.1 Estimation of severity distribution

In order to select all possible distributions that best fit observed data it is necessary to analyze the features of an operational loss database. Most empirical distribution show a positive skewness and high kurtosis. It happens because generally an operational loss database is characterized by many low intensity losses and few high intensity losses or extreme losses. In this regard, the most commonly used distribution is the Log-normal, although there exist a quite list of distributions that can be used in modelling the severity of operational losses like Weibull, Exponential, Gamma etc. (Dahen and Dionne, 2010; Klugman et al., 2012). The main issue is related to the fact that they tend to underestimate the losses in the right tail of the distribution. In order to avoid such problem, it is common practice to split up the monetary distribution impact in two parts. One for the body and one for the tail. They are then aggregated through a mixture function to obtain a single model. Such a new distribution will result more reliable in taking into account the impact of rare events. Crucial to this process is the threshold "*u"*, which will be used as discriminatory element to separate the body from the tail of the distribution. Therefore, the severity distribution function of a single internal operational loss is a mixture between two different distributions (Figure 2).

Figure 2. Mixture function



In addition, a bank may not keep recording of very low impact losses. In this case the body distribution must be modified in such a way to take into account the truncation effect of data below a threshold *T*. With this regard it is necessary to modify the body distribution function, and introducing the left truncated conditional probability distribution. The most widely used approaches to estimate parameters are the followings: a system of equations equal to the number of parameters (the popular methods are: moments, percentile matching, probability weighted moments) and an optimization process. As far as tail parameters estimation concerns, Extreme Value Theory (EVT) provides a theoretical framework for studying rare event by focusing on the tails of the probability distributions. Two different approaches are used to estimate the tail distribution when dealing with EVT: Block Maxima and Peaks Over Threshold methodology (Abbate et al., 2009; Cruz, 2002; Da Costa Lewis, 2004).

4.2 Estimation of frequency distribution

The estimation of frequency distribution of operational loss implies the representation of the pattern of observed operational events through a discrete random variable. The most commonly used distribution is the Poisson (Table 1).

Table 1. Poisson distribution

Distributions		Parameters
Poisson	$p(x) = \lambda^x \frac{\exp(-\lambda)}{x!}$	$(\lambda > 0)$

When a bank decides not to account operational losses below a certain threshold *T*, it is necessary to estimate the parameter λ taking into account the truncation effect:

$$\hat{\lambda} = \frac{\hat{\lambda}_s}{1 - F_b(T;\theta)}$$

Where $\hat{\lambda}_s$ is the estimated parameter from the loss events database. After having computed the

parameter, it is possible to estimate the parameter for the body and the tail of the distribution.

4.3 Aggregation of loss distribution

Having separately assessed both frequency and severity distributions, it is now necessary to combine them into one aggregate loss distribution. The most used aggregation methodology is Monte Carlo simulation. However, before moving on the simulation itself we need to satisfy the independence hypothesis between frequency and severity in order to adopt that methodology. The aggregate loss distribution comes from a convolution process between the previous estimated frequency and severity distributions. As well as in the previous stage we need to carry out the Monte Carlo simulation separately for the body and tail. After that, through a convolution process we can join them to estimate the aggregate loss distribution.

The Monte Carlo simulation involves different steps for the body and tail of the loss distribution: sampling the number of annual losses, generating as many uniform random variables as demanded by the frequency, using such variables as probabilities to find out the quantile in the chosen severity distribution function. After repeating several times and sorting out the loss data from the smallest to the largest, the aggregate loss distribution is obtained. Finally, the aggregate loss distribution for the body and the tail are summed to obtain the annual aggregate loss distribution.

4.4 Aggregation of risk classes and estimation of capital-at-risk

The most conservative approach requires to estimate the total capital as the sum of the capital-at-risk of each risk class. This approach assumes a perfect linear correlation hypothesis among different risk classes. To remove such limitation and estimate the overall annual loss distribution may be used an aggregating technique based on a Copula methodology.

A Copula distribution function is obtained by starting from marginal distributions and dependence structure. The main dependence measures between random variables are the followings: the Pearson linear correlation, the rank correlation coefficients, and tail dependence. The main Copula functions exploited for their technical prescriptions in the operational risk framework are the Archimedean Copula and the Elliptical Copula functions (Chernobai et al., 2007; McNeil et al., 2005). Once identified the most suitable copula that represents the operational loss multivariate distribution, then it is possible to determine the capital requirement using Value-at-Risk measurements.

5 A simulated operational risk transfer strategy

This section of the paper aims to demonstrate how an operational risk transfer policy based on insurance contracts can mitigate the impact on the bank regulatory capital. The analysis is based on simulated data instead of empirical ones because of the high sensitivity and confidentiality of banks' operational loss databases. The simulation is conducted on the Loss Data Collection Exercise that has been carried out by the Operational Risk Subgroup of the Standards Implementation Group (SIGOR). The analysis is performed using the statistical language R. The analysis is carried out in two parallel steps. In the first one, after a comprehensive description of the datasets, the operational risk capital requirement is estimated. In the second one, the mitigating impact on the regulatory capital is the result of a transferring operational risk policy.

In order to carry out the analysis, two aspects are crucial. Firstly, the definition of operational risk class in order to satisfy the hypotheses of independence and identical distribution. Secondly, the number of risk classes to be considered in the analysis. As regards the number of risk classes, the low amount of data stored within the database would not provide full robustness of the statistical results. Thus, it would be not possible to use as risk class the minimum one – i.e. intersection of business line and event type – because almost each bank lies at the initial stage in the use of such methodology. In addition it will be used the breakdown by event type rather than business line, since the former provides a direct insight into transferring techniques, and therefore the effects on regulatory capital (Cruz, 2002; Davis, 2006).

In particular, the data on which the analysis will be performed are extracted from a simulated operational losses database. Such database will represent the internal operational database that a bank may hold. Moreover, in order to get closer to the reality it is assumed that a bank does not keep recording of losses under 2,000 euro. The features of such database are summarized in Table 2 and Table 3.

5.1 Modelling the severity distribution

Earlier we explained the reason why we need to spilt up the severity distribution in two parts. One for the body and the other for the tail. In addition, we need to take into consideration the truncation effect since we are assuming a bank does not keep recording for losses under 2,000 euro. Over the course of this paper we present the analysis exclusively for the Event Type 1, as we can easily extend the same conclusions to the other risk classes. Firstly, for each risk class we need to justify the use of Extreme value theory and demonstrate that loss data satisfies the hypothesis of independence and identical distribution (Abbate et al., 2009; Klugman et al., 2012).

Table 2. Loss frequencies by event type

Code Event Type	Event Type	Operational Events within one year
Et1	Internal Fraud	1374
Et2	External Fraud	8564
Et3	Employment. Practices. & Workplace. Safety	5714
Et4	Clients Products & Business Practices	5915
Et5	Damage to Physical Assets	383
Et6	Business Disruption & System Failures	642
Et7	Execution, Delivery & Process Management	9970
	All	**32562**

Table 3. Loss severity by event type

Event Type	Minimum Loss	Maximum Loss	Median Loss	Mean Loss	Standard Deviation	Skewness	Kurtosis
Et1	2,058	398,6143	33,228	85,065	186,318	9.8	159
Et2	2,006	4,089,191	23,861	5,2347	109,477	11.8	282
Et3	2,000	3,424,911	25,816	71,738	159,956	7.8	97
Et4	2,021	343,170,300	293,599	1,548,775	7,400,414	24	887
Et5	2,023	375,254	12,177	25,469	37,946	4.38	27
Et6	2,119	765,084	10,564	24,754	50,163	8,28	96
Et7	2,094	42,775,230	146,408	416,194	1,070,166	14	381

Figure 3. Event type 1 - Box plot (left) and log-scale box plot (right)

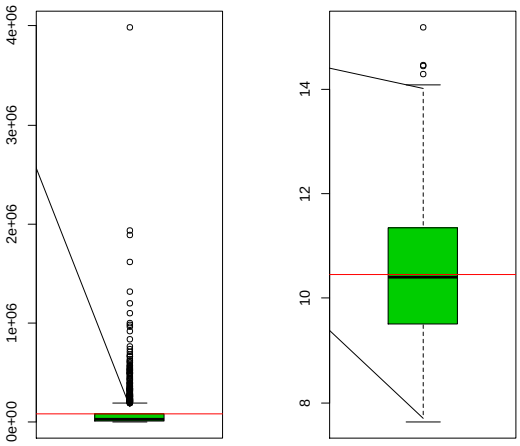
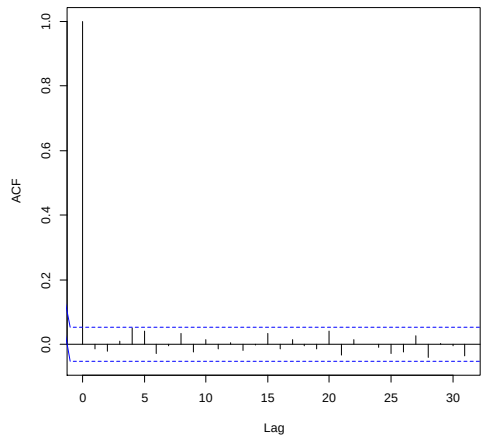


Figure 4. Autocorrelation plot

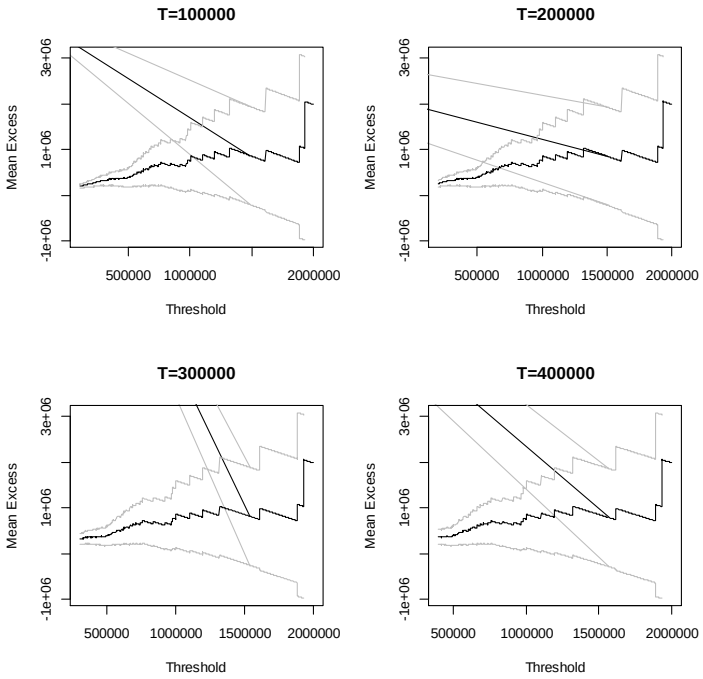


From the box-plot (Figure 3) we can immediately observe that data are showing a highly skewed to the right such that to justify the use of Extreme value theory. In order to check the independence and identical distribution hypothesis we have to look at the autocorrelation plot and to the Box-Jenkins test.

Assuming that the sample data is sorted by accounting date, the autocorrelation plot (Figure 4) shows that the independency hypothesis is satisfied since the autocorrelation values are within the 95% confidence level. Whereas, Box-Jenkins test returns a statistic value X-squared = 0.277 with consequent p-value=0.5987. Therefore, we can apply the methodology that has been introduced so far.

Now, it is possible to move onto the parameters estimation for the body distribution. In order to explain the procedure and make it simple, we choice to fit our simulated internal data to two theoretical distributions, the left-truncated log-normal distribution and the left-truncated Weibull distribution. Before moving on the parameters estimation itself, we need to identify the threshold over which the tail has to be estimated through extreme value theory. In particular we decided to recur on Peaks Over Threshold methodology, in order to find out the generalized Pareto distribution which explains the tail behavior. As mentioned earlier a primary tool is the Sample mean excess plot.

Figure 5. Sample mean excess plot for different level of T



The four graphs pictured in Figure 5 represent the sample mean excess function for different threshold levels. If the Generalized Pareto distribution is a good fit to the tail, the plot should become approximately linear. Our purpose here is to pick the largest threshold beyond which the plot starts to become linear. Indeed, if the threshold is chosen too high, then there are not enough exceedance over the threshold to obtain good estimators of the extreme value parameters, and consequently, the variances of the estimators are high. Conversely, if the threshold is too low, the Generalized Pareto distribution may not be a good fit to the excesses over the threshold and there will be bias in the estimations. Other useful tools are the parameter stability plots (Figure 6), which help us to pick a right threshold. They should become stable above the right threshold (Horbenko et al., 2011).

In our example a threshold of T = 400,000 seems good enough both in terms of linearity of the sample mean excess and stability of the parameter estimates.

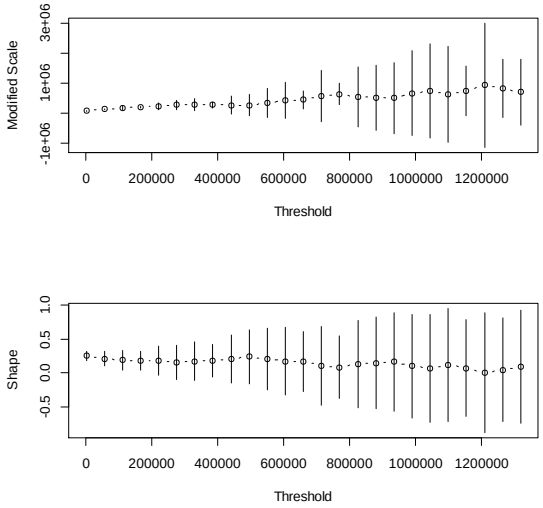
Once we have identified the right threshold we can estimate the parameters from the left-truncated log-normal distribution. So, firstly we need to find out the log-likelihood for a left-truncated log-normal distribution. In general we define the likelihood of a particular model the following expression:

$$L(x;\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta})$$

Where the maximum likelihood estimate is:

$$\widehat{\boldsymbol{\theta}} = argmax\ L(x;\boldsymbol{\theta});\ \boldsymbol{\theta} \in \boldsymbol{\Theta 0}$$

Figure 6. Parameter stability plots



Actually, it is more convenient to deal with the log-likelihood, specified as:

$$l(x;\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i;\boldsymbol{\theta})$$

Maximizing the log-likelihood first requires taking the partial derivatives with respect to the parameters and setting them equal to zero:

$$\frac{\vartheta\, l(x;\boldsymbol{\theta})}{\vartheta\theta} = 0$$

Sometimes it is not possible to get an explicit expression, in those cases we recurr to numerical optimization methods. Comig back to the issue regarding to left-truncated distributions we need to set up the likelihhod function of a conditional density function, as in the following expression:

$$L(x;\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{f_b(x;\theta)}{1 - F_b(T;\theta)}$$

With corresponding log-likelihhod function:

$$l(x;\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \frac{f_b(x;\boldsymbol{\theta})}{1 - F_b(T;\boldsymbol{\theta})} = \sum_{i=1}^{n} \log f_b(x;\boldsymbol{\theta}) - n \log\big(1 - F_b(T;\boldsymbol{\theta})\big)$$

Firstly, we proceed with fitting our data sample with the left-truncated log-normal distribution. As regards to the left-truncated log-normal, we cannot obtain an explicit expression for the MLE estimate. In such a case we decide to perform the estimate recurring to the Nelder-Mead numerical optimization method in order to find out the two parameters characterizing the distribution. We report the R code to process the MLE estimates:

ltlnorm<-function(x,meanlog,sdlog)

dlnorm(x, meanlog, sdlog) / plnorm(2000, meanlog, sdlog,lower.tail=FALSE)

fitdistr(x,ltlnorm,start=list(meanlog=5,sdlog=2))

The resulting estimates are:

$$\hat{\mu} = 10.38 \;\; \hat{\sigma} = 1.38$$

In order to check how well our model fits a set of observations, we perform both graphical and quantitative tests. Firstly, we report the Q-Q plot relative to the body of the severity (Figure 7).

As the picture shows the model seems to fit well the observations especially for values under the threshold *u* (green line), while for values above *u* the model seems to lose adaptabilty to the data. This is not a problem since those observations will be processed and modeled in the tail of the distribution. Regarding to the quantitative tests, we report the well studied Kolmogorov-Smirnov test and Anderson-Darling test for left-truncated data (Table 4).

All the p-values are sufficiently high. So, the null hypothesis is not rejected. The same analysis has been carried out assuming as theoretical distribution a left-truncated Weibull distribution.

The resulting MLE estimates are:

$$\hat{\alpha} = 7.33 \quad \hat{\beta} = 10.85$$

We report the Q-Q plot resulting from these estimates (Figure 8).

Figure 7. Q-Q plot for left truncated Log-Normal distribution

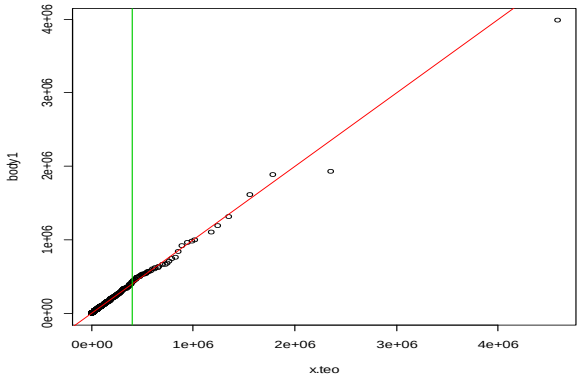
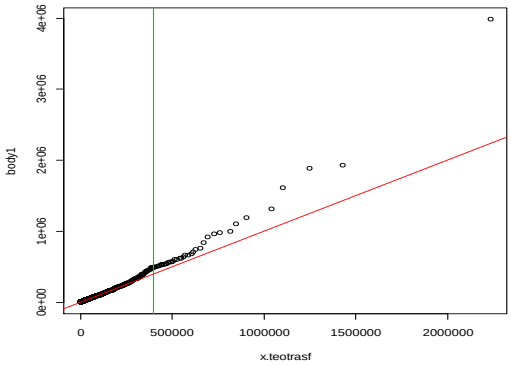


Table 4. Goodness test results between the left-truncated log-normal distribution and the sample data

Test	Statistics	p-value
Kolmogorov-Smirnov	0.7335	0.185
Anderson-Darling	2.0298	0.618
Anderson-Darling up	71.9778	0.346

Figure 8. Q-Q plot for left truncated Weibull distribution



As showed by the picture, the model seems to fit well the observations especially for values under the threshold *u* (green line), while for values above *u* the model seems to lose adaptabilty to the data, as well as the previous case. Regarding to the quantitative tests, again we report the Kolmogorov-Smirnov test and Anderson-Darling test for left-truncated data (Table 5).

Table 5. Goodness test results between the left-truncated Weibull distribution and the sample data

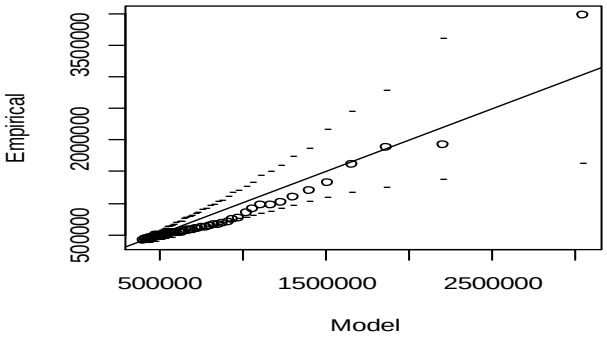
Test	Statistics	p-value
Kolmogorov-Smirnov	1.13	0.0099
Anderson-Darling	9.52	0.0222
Anderson-Darling up	3397.245	0.009259

All the quantitative tests do not reject the null hypothesis. Therefore the body of the severity distribution relating to Event type 1 will be modeled as a log-normal distribution. Once identified the body distribution we can move forward to the estimation of the tail distribution. Relating to the parameter estimates, we apply again maximum likelihood estimation. The resulting scale and shape parameters are:

$$\hat{\sigma} = 368{,}400 \quad \hat{\varepsilon} = 0.181$$

We report the Q-Q plot for the tail distribution (Figure 9).

Figure 9. Q-Q plot for generalized Pareto distribution



As the picture shows the model seems to fit well the observations for values over the threshold u.

Now for any event type we report the model and the corresponding parameter estimates (Table 6).

Table 6. Body and tail severity distributions parameter estimates

Event Type	Model Body	$\hat{\mu}$	$\hat{\sigma}$	Boundary threshold (u)	Model tail	$\hat{\sigma}(scale)$	$\hat{\varepsilon}(shape)$
Et1	Lognormal	10.38	1.38	400,000	GPD	368,400	0.181
Et2	Lognormal	10.04	1.24	500,000	GPD	385,600	0.111
Et3	Lognormal	10.05	1.48	600,000	GPD	439,600	0.084
Et4	Lognormal	12,58	1.80	5,000,000	GPD	1,130,000	0.279
Et5	Lognormal	9.38	1.18	50,000	GPD	50,030	0.104
Et6	Lognormal	8.96	1,41	100,000	GPD	115,600	0.183
Et7	Lognormal	11.87	1.46	5,000,000	GPD	4,114,000	0.1054

5.2 Modelling the frequency distribution

The Poisson distribution represents the frequency distribution of an operational event. The only parameter λ is estimated through the method of moments (Horbenko et al., 2011). We only use the annual frequency of occurrence for operational losses higher than the threshold T = 2,000 euro.

$$\hat{\lambda}sample = \frac{\sum_{i=1}^{365} n_i}{365} = \frac{1374}{365} = 3.76$$

This value does not take into account the effect of operational events causing a loss below a threshold T. Thus, in order to deal with a more reliable estimate we have to consider the truncation effect:

$$\hat{\lambda} = \frac{\hat{\lambda}sample}{P_{body}(loss>2000)} = \frac{3.76}{1-P(2000)} = \frac{3.76}{0.9779876} = 3.85$$

Starting from this estimation, we can define the frequency of loss for the body and the frequency of occurrence for the tail, as it follows:

$$P_{body}(400000) = 0.9659312 \quad P_{body}(2000) = 0.02201243$$

$$\hat{\lambda}_{body} = \hat{\lambda}\left[P_{body}(400000) - P_{body}(2000)\right] = 3.85(0.9659 - 0.02201) = 3.63$$

$$\hat{\lambda}_{tail} = \hat{\lambda}\left[1 - P_{body}(400000)\right] = 3.85(1 - 0.9659) = 0.13$$

Now for any event type we report the corresponding parameter estimates (Table 7).

Table 7. Body and tail frequency distributions parameter estimates

Event Type	Model	$\hat{\lambda}_{body}$	$\hat{\lambda}_{tail}$
Et1	Poisson	3.85	0.13
Et2	Poisson	23.29	0.16
Et3	Poisson	15.34	0.24
Et4	Poisson	15.25	0.95
Et5	Poisson	0.91	0.13
Et6	Poisson	1.69	0.08
Et7	Poisson	27.09	0.22

5.3 Severity and frequency convolution

After estimating the severity and frequency distribution, the next step involves their aggregation in order to perform the aggregate loss distribution. Such aggregating procedure is based on a convolution process. It is a mathematical operation through which two functions return a third function (Da Costa Lewis, 2004; Frachot et al., 2001; Shevchenko, 2010). Assuming that the two distributions satisfy the independence hypothesis necessary to exploit the convolution mechanism, it is therefore possible to implement the algorithm based on Monte Carlo simulation. As mentioned, such algorithm provides subsequent sampling from the frequency and severity distribution to return the aggregate loss distribution. Once again we report the analysis only for Event type 1. In details, we run the simulation 10^6 times to obtain a suitable distribution.

Let i = 1…… *N*, the Monte Carlo simulation methodology involves the following steps:

a. Sampling the number of annual losses for the body:

$$m_i \sim Pois(\hat{\lambda}_{1b} = 3.63)$$

b. Generating as many uniform random variables as demanded by the frequency.

c. Those variables will be used as the probability (*p*) that we use to find out the quantile in the chosen body severity distribution function.

$$b_{i,r} \sim F_{1b}(x; (\hat{\mu} = 10.38; \hat{\sigma} = 1.38/x \geq 2000))$$

With $r = 1, \ldots\ldots, m_i$

d. Repeating several times the process to obtain the aggregate loss distribution for the body.

In the same way we proceed to get the aggregate loss distribution for the tail:

a. Sampling the number of annual losses for the tail:

$$n_i \sim Pois(\hat{\lambda}_{1t} = 0.13)$$

b. Generating as many uniform random variables as demanded by the frequency.

c. Those variables will be used as probabilities (*p*) that we use to find out the quantile in the chosen tail severity distribution function:

$$t_{i,r} \sim F_{1t}(x; (\hat{\sigma} = 368{,}400; \hat{\varepsilon} = 0.181))$$

With $r = 1, \ldots\ldots, n_i$

d. Repeating several times the process to obtain the aggregate loss distribution for the tail.

Lastly, in order to obtain the annual aggregate loss distribution for a certain risk class we compute it, as it follows:

$$G_i = \sum_{r=1}^{m_i} b_{i,r} + \sum_{r=1}^{n_i} t_{i,r} + c$$

Where *c* is the mean of the operational loss data empirical distribution below a certain threshold *T*.

Once produced seven aggregate simulated loss distributions, we can as well as extract the expected loss and the Value-at-Risk, with a 99,9% confidence level, and assuming to use the conservative approach, compute the capital at risk for the bank as a whole (Table 8).

Table 8. Value-at-risk

Code Event Type	Expected Loss €	Unexpected Loss €	Value-at-Risk 99,9% €
Et1	361,728	4,379,121	4,740,849
Et2	1,233,710	3,675,625	4,909,335
Et3	1,186,874	6,996,498	8,183,372
Et4	24,360,690	297,000,914	321,361,604
Et5	29,027	489,534	518,561
Et6	47,414	1,001,130	1,048,545
Et7	12,442,230	52,922,301	65,364,531
Capital-at-Risk	**€ 39,661,674**	**€ 366,465,123**	**€ 406,126,797**

In the graphs below (Figure 10) we report the histograms of the square root transformations of the seven aggregate annual loss distributions. These graphs show the existence of a strong lack of homogeneity among different event types, pointing out that each event type is characterized by different risk drivers. Heterogeneity, which is explained by the different impact that each individual operational event has in terms of operational loss. Moreover, we can immediately observe, as well as which types of events have a greater weight in determining the capital at risk, which ones represent major concerns for the operational risk manager and the bank as a whole. In particular, the graphs show that the event type 4 represents the major bank's source of operating losses, given that approximately 18% of operational events explains more than 79% of regulatory capital.

Figure 10. Histograms for the sqrt transformation of the seven event type distribution

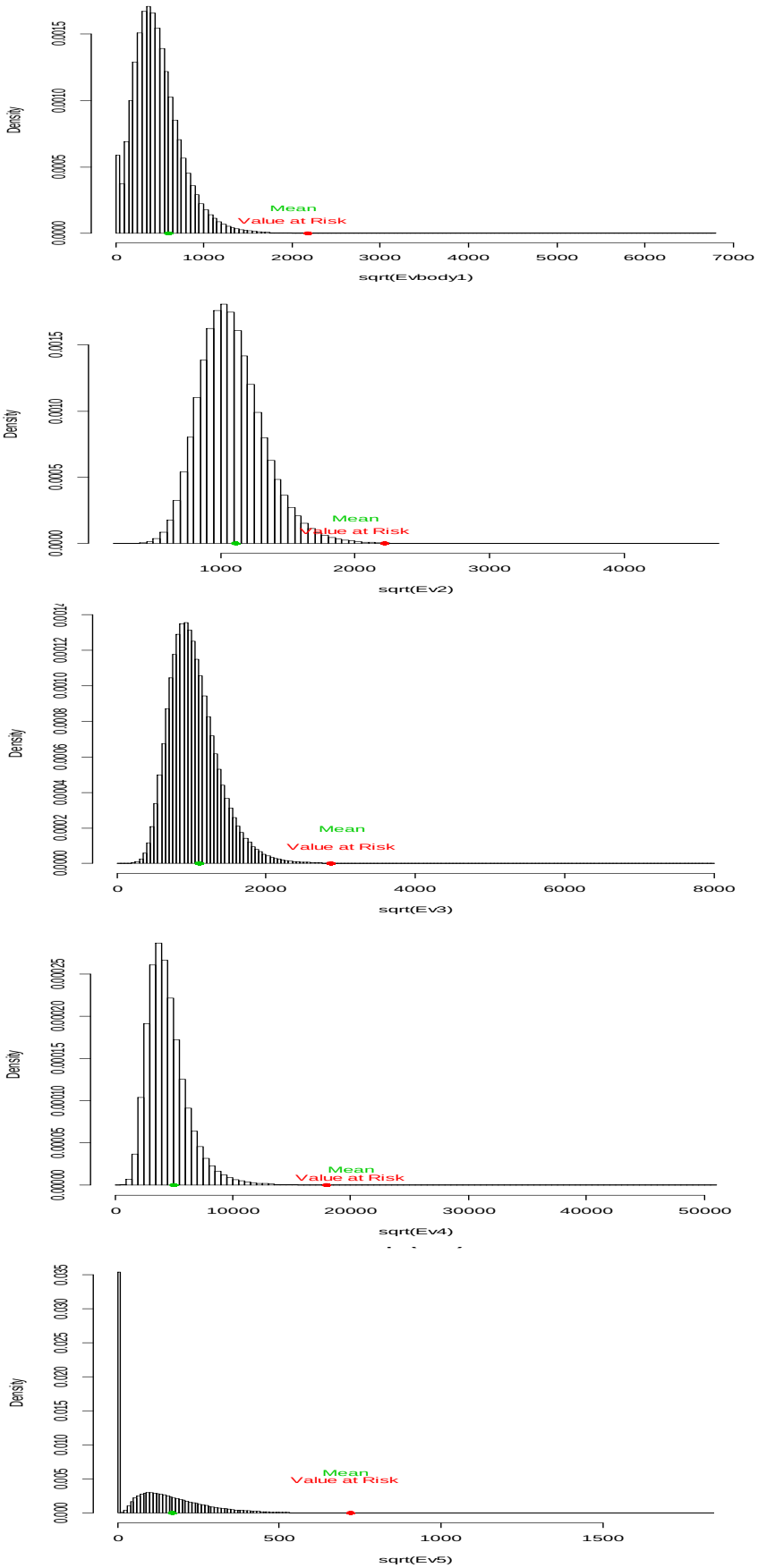
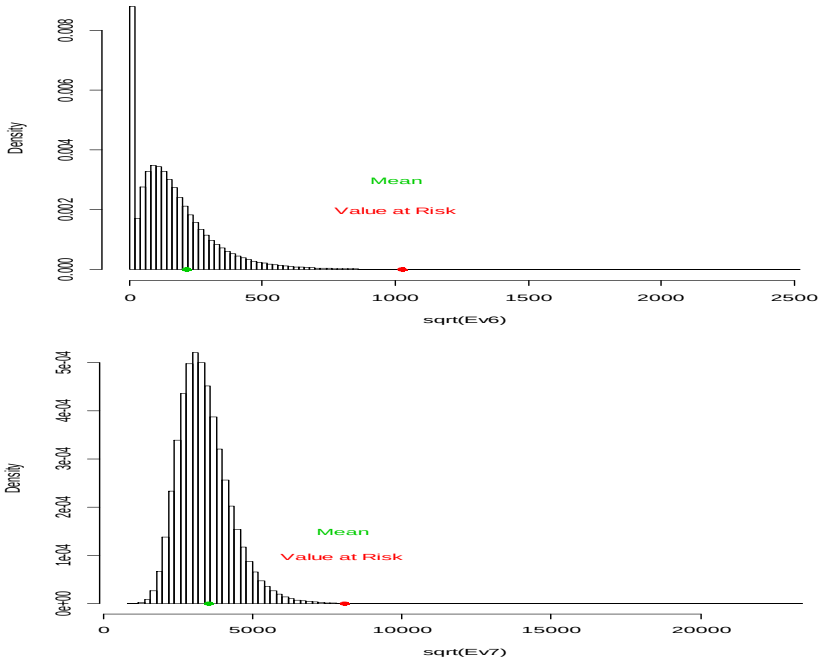


Figure 10. Histograms for the sqrt transformation of the seven event type distribution (continued)



5.4 The overall annual loss distribution

The sum of 7 value-at-risk measures that has been calculated on the annual loss distributions to determine the bank's capital at risk involves the implementation of an extremely protective policy in terms of regulatory capital. In fact, we are implicitly assuming that there is a perfect correlation among operational loss distributions. If we remove such hypothesis it will be necessary to carry on with the overall annual loss distribution. Consequently, we can estimate the value-at-risk with a 99,9% confidence level. Such solution leads to a more appropriate estimation of the capital-at-risk in banking.

In this regard, we carry out the analysis using the elliptical copula family, since it allows to take into consideration the real dependency structure among different event types and to mark the role of the dependence structure in the proximity of extreme values.

$$R = \begin{pmatrix} & Et1 & Et2 & Et3 & Et4 & Et5 & Et6 & Et7 \\ Et1 & 1 & 0.00022 & -0.00019 & -0.0016 & 0.00051 & -0.00048 & -0.00087 \\ Et2 & & 1 & -0.00005 & 0.00076 & -0.00053 & -0.00188 & -0.00119 \\ Et3 & & & 1 & 0.00089 & 0.00039 & 0.00118 & -0.00128 \\ Et4 & & & & 1 & -0.00154 & -0.00025 & 0.00054 \\ Et5 & & & & & 1 & -0.00040 & 0.00061 \\ Et6 & & & & & & 1 & -0.00020 \\ Et7 & & & & & & & 1 \end{pmatrix}$$

We estimate the overall annual loss distribution with the Monte Carlo simulation method. Then we compare the results of Gaussian copula with those of t-Student copula. For the latter, it is necessary to estimate the parameter *v*, through the use of the maximum likelihood estimator.

The algorithm for both distributions has been carried out for a number of simulations equal to 10^6. In the following figures, we report the square root transformation of the histograms of the overall annual loss distribution in three cases: the existence of a perfect positive correlation among all risk classes, the Gaussian copula, and the t-Student copula (Figure 11).

The results are summarized in the following Table 9.

The above mentioned results confirm two important conclusions. The first one refers to a substantial saving capital requirement by implementing an aggregation mechanism, which takes into account the correlation among different risk classes. We obtain a lower regulatory capital when we use the Gaussian or t-Student copula in comparison to the sum of single value-at-risk for each annual loss distribution. The second one refers to the tail dependence of the t-Student copula. Such peculiarity brings to a higher regulatory capital of the t-Student copula in comparison to the Gaussian copula one.

Figure 11. Histograms for the sqrt transformation overall annual loss distribution

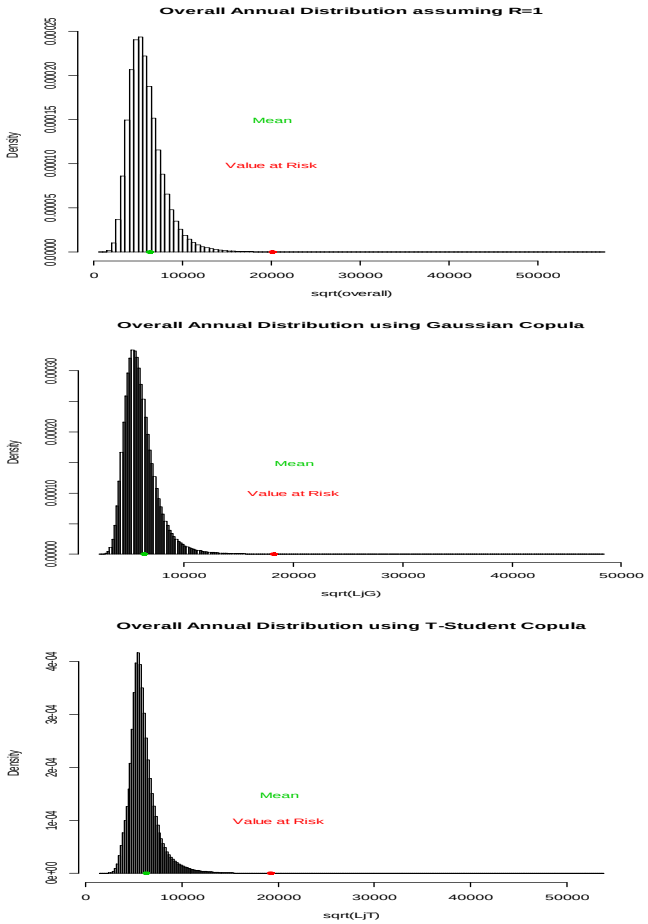


Table 9. Value-at-risk using different aggregating mechanisms

Aggregating Mechanism	Expected Loss €	Unexpected Loss €	Value-at-Risk 99,9% €
R=1	39,661,674	366,465,123	406,126,797
Gaussian Copula	39,666,040	290,961,758	330,627,798
t-Student Copula	39,641,997	328,155,528	367,797,525

5.5 Hedging against operational risk

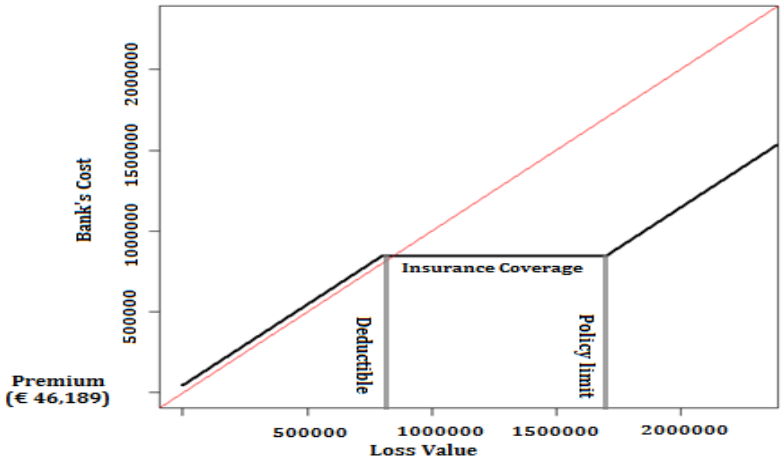
This section of the paper aims to illustrate how an effective operational risk transfer strategy can result in saving regulatory capital. In order to analyse the risk mitigating strategy, let us suppose a bank decides to negotiate 5 insurance contracts that aim to attenuate the negative impact of operational risk. In particular, the structure of each insurance contract has been set to avoid moral hazard problems. For this purpose, insurance contracts fix deductibles (amounts of money subtracted from the value of a loss, which is not covered by insurance) and policy limits, in such a way to not encourage bank's opportunistic behavior. With reference to the insurance pricing, we consider the pure risk premium. The Table 10 shows some details of 5 insurance contracts.

Table 10. Insurance contracts: some details

Code Event Type	Premium €	Deductible €	Policy Limit €
Et-1	46,189	800,000	1,700,00
Et-2	8,930	2,000,000	3,800,000
Et-3	22,725	3,000,000	5,800,00
Et-4	1,680,030	55,000,000	95,000,00
Et-7	237,234	20,000,000	40,000,00
	€1,995,108		

In order to show how insurance contracts work, it is necessary to observe the following graph, where we report the Event type 1 annual loss distribution in the presence of an insurance contract that has the above mentioned characteristics (Figure 12).

Figure 12. Event type 1 - Insurance coverage effect



The insurance contract transforms the distribution in a step function that is characterized by the followings:

$$\begin{cases} P + ET_1(x), & x \le D_1 \\ D_1, & D_1 \le x \le D_2 \\ ET_1(x) - D_1, & x > D_2 \end{cases}$$

with ET_1 the loss value, and D_1 and D_2 the value of deductible and policy limit respectively.

Briefly, the insurance contract modifies the monetary impact coming from an operational event. Thus, we need to modify the severity distribution (Banks, 2004; Committee of European Banking Supervisors, 2009; Cruz, 2002). Conversely, if insurance contracts have not any effects on the occurrence of an operational event we will not need to modify the previously estimated frequency distribution.

The measurement of the new capital requirement takes into account the mitigating effect of insurance contracts. The estimation of the overall annual loss distribution has been carried out using Monte Carlo simulation and copula methodologies. The following pictures report the histograms of the overall annual loss distribution, respectively for the Gaussian and t-Student copula aggregating mechanisms (Figure 13, 14).

Figure 13. Histograms for the sqrt transformation overall annual loss distribution with insurance contracts (using Gaussian Copula)

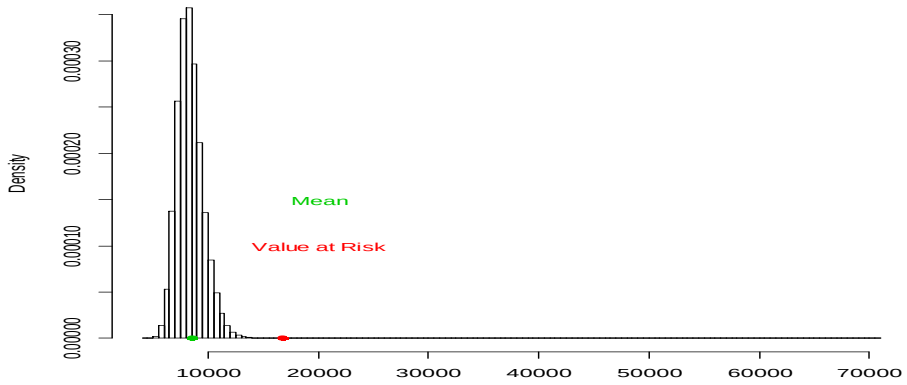
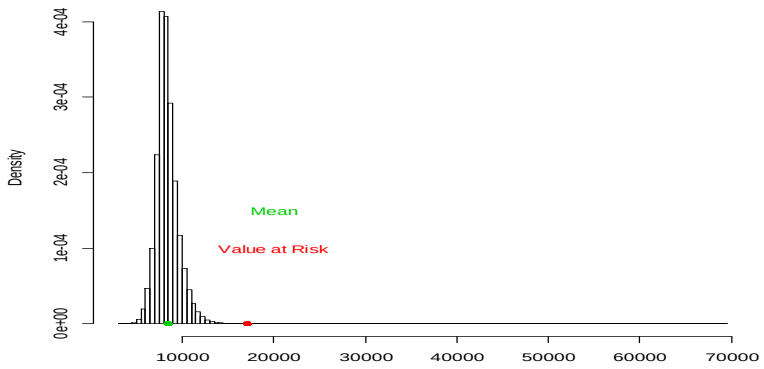


Figure 14. Histograms for the sqrt transformation overall annual loss distribution with insurance contracts (using t-Student Copula)



Finally, in order to summarize the results of an operational risk transfer strategy, the Tables 11 and 12 show the estimation of value-at-risk using the different aggregation mechanisms (under insurance contracts) and a comparison between regulatory capital with insurance contracts and regulatory capital without insurance contracts.

Table 11. Value-at-risk under insurance contracts (using different aggregation mechanisms)

Aggregating Mechanism	Expected Loss €	Unexpected Loss €	Value-at-Risk 99,9% €
Gaussian Copula	72,218,751	206,547,581	278,766,332
t-Student Copula	72,172,803	218,604,831	290,777,634

Table 12. A comparison of capital requirements with and without insurance contracts

Aggregating Mechanism	Var 99% without insurance contracts	Var 99,9% with insurance contracts	% Saving on Capital Requirement
Gaussian	330,627,798	278,766,332	16%
t-Student	367,797,525	290,777,634	21%

The operational risk transfer strategy involves a regulatory capital saving to an extent of 16% when we use a Gaussian copula and 21% when we use a t-Student copula. Although the capital adequacy regulation imposes a maximum level (20%) of the mitigating effect on the capital requirement (Basel Committee on Banking Supervision, 2006, 2010). Therefore, with the adoption of t-Student copula methodology the regulatory capital should not be less than the 80% of the value-at-risk without insurance contracts.

6 Conclusion

Banking industry has made significant progress over the past years in understanding, measuring and managing operational risk. Banking authorities have been pressuring banks to adopt a proactive operational risk management. In addition they have imposed protective measures based on the provision of a minimum level of regulatory capital to absorb risk operational losses.

The paper has been designed to demonstrate how an effective operational risk management provides a regulatory capital saving, and a resulting reduction of bank capital costs. A simulated operational losses database supported the operational risk transfer strategy. The estimation of the overall annual loss distribution has been carried out using Monte Carlo simulation and copula methodologies. The operational risk transfer strategy involves a regulatory capital saving to an extent of 16% when we use a Gaussian copula and 21% when we use a t-Student copula.

In addition, it is important to note several aspects that need further developments. Firstly, nowadays only a limited number of banks are using an advanced methodology to estimate the regulatory capital. Nevertheless, it is only through the use of advanced measurement tools that it is possible to implement an effective hedging strategy. Therefore, in a perspective of proactive risk management, it is necessary to stimulate the adoption of AMA methodologies. Sources of incentives may arise from the standardization of the methodologies to estimate the regulatory capital, and the increasing accessibility of AMA methodologies to smaller banks.

Finally, in order to avoid that the operational risk transfer market becomes a possible source of financial instability in the banking industry it is necessary to develop transparent procedures and policies, to ensure

that a certain amount of operational risk still stays within the responsibilities of the bank management.

References

1. Abbate, D., Farkas, W. and Gourier E. (2009), "Operational Risk Quantification using Extreme Value Theory and Copulas: from Theory to Practice", Journal of Operational Risk, Vol. 3 No. 1, pp.1-24.
2. Alexander, C. (2003), Operational Risk: Regulation, Analysis and Management, Prentice Hall, London.
3. Aprile, G. A. (2007), "L'importanza dei dati esterni nella misurazione dei rischi operativi", Bancaria, Vol. 35, No. 4, pp.25-48.
4. Banks, E. (2004), Alternative Risk Transfer. Integrated Risk Management through Insurance, Reinsurance and the Capital Markets, John Wiley & Sons, New Jersey.
5. Basel Committee on Banking Supervision (2001), Working Paper on the Regulatory Treatment of Operational Risk, September.
6. Basel Committee on Banking Supervision (2003), Operational Risk Transfer across Financial Sectors, August.
7. Basel Committee on Banking Supervision (2006), International Convergence of Capital Measurement and Capital Standards. A Revised Framework. Comprehensive Version, June.
8. Basel Committee on Banking Supervision (2010), Recognising the risk-mitigation impact of insurance in operational risk modelling, October.
9. Basel Committee on Banking Supervision (2011a), Principles for the Sound Management of Operational Risk, June.
10. Basel Committee on Banking Supervision (2011b), Operational Risk. Supervisory Guidelines for the Advanced Measurement Approaches, June.
11. Bessis, J. (2009), Risk Management in Banking, John Wiley & Sons, New York.
12. Birindelli, G. and Ferretti, P. (2006), Evoluzione del rischio operativo nelle imprese bancarie, Giuffrè, Milan.
13. Birindelli, G. and Ferretti, P. (2009), Il rischio operativo nelle banche italiane. Modelli di gestione e disclosure, Bancaria Editrice, Rome.
14. Chernobai, A. S., Rachev, S. T. and Fabozzi, F. J. (2007). Operational risk. A Guide to Basel II Capital Requirements, Models, and Analysis, John Wiley & Sons, New Jersey.
15. Committee of European Banking Supervisors (2009), Guidelines on Operational Risk Mitigation Techniques, December.
16. Cosma, S. (2008), La misurazione del rischio operativo nelle banche, Bancaria Editrice, Rome.
17. Cruz, M. G. (2002), Modelling, Measuring and Hedging Operational Risk, John Wiley & Sons, New York.
18. Dahen, H. and Dionne, G. (2010), "Scaling Models for the Severity and Frequency of External Operational Loss Data", The Journal of Banking and Finance, Vol. 34 No. 3, pp.1484-1496.
19. Davis, E. (2006), The Advanced Measurement Approach to Operational Risk, Risk Books, London.
20. Da Costa Lewis, N. (2004), Operational Risk with Excel and VBA: Applied Statistical Methods for Risk Management, John Wiley & Sons, New Jersey.
21. Frachot, A., Georges, P. and Roncalli, T. (2001), "Loss Distribution Approach for operational risk, Groupe de Recherche Opérationnelle Crédit Lyonnais", available online at the following URL: http://thierrroncalli.com/download/lda.pdf. Accessed 20 January 2015.
22. Gabbi, G., Marsella, M. and Massacesi, M. (2005), Il rischio operativo nelle banche, Egea, Milan.
23. Horbenko, N., Ruckdeschel, P. and Bae, T. (2011), "Robust Estimation of Operational Risk", The Journal of Operational Risk, Vol. 6 No. 2, pp.3-30.
24. Hull, J. (2012), Risk Management and Financial Institutions, John Wiley & Sons, New York.
25. Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012), Loss Models from Data to Decisions, John Wiley & Sons, New Jersey.
26. McNeil, A. J., Frey, R. and Embrechts, P. (2005), Quantitative Risk Management: Concepts, Techniques and Tools, Princeton University Press, Princeton.
27. Resti, A. and Sironi, A. (2007), Risk Management And Shareholders' Value In Banking: From Risk Measurement Models To Capital Allocation Policies, John Wiley & Sons, London.
28. Resti, A. and Sironi, A. (2008), Rischio e valore nelle banche, Egea, Milan.
29. Santomero, A. and Babbel, D. (2012), Financial Market, Instruments, and Institutions, McGraw-Hill, New York.
30. Scannella, E. (2005), "Alcune osservazioni sul processo di revisione dei requisiti minimi di patrimonializzazione delle aziende di credito", Rivista Italiana di Ragioneria e di Economia Aziendale, Vol. 104 No.3-4, pp.154-167.
31. Shevchenko, P. V. (2010), "Calculation of Aggregate Loss Distributions", The Journal of Operational Risk, Vol. 24 No. 5, pp.3-40.
32. Sironi, A. (2003), "Il rischio operativo: una nuova sfida per le banche italiane", Economia & Management, Vol. 30, No. 1, pp.1-15.
33. Tutino, F, Birindelli, G. and Ferretti, P. (2011), Basilea 3. Gli impatti sulle banche, Egea, Milan.
34. Tutino, F, Birindelli, G. and Ferretti, P. (2012), "Bank Capital and Basel 3. Impacts on Italian Banks", Corporate Ownership & Control, Vol. 10 No. 1, pp.75-87.

SUBSCRIPTION FORM
TO "CORPORATE OWNERSHIP & CONTROL", "CORPORATE BOARD: ROLE, DUTIES & COMPOSITION", "RISK GOVERNANCE & CONTROL: FINANCIAL MARKETS & INSTITUTIONS", "JOURNAL OF GOVERNANCE AND REGULATION"

Copy this form and follow guidelines to fill it up. I would like to buy (underline what you choose):

For individual subscribers:

1. Journal of Corporate Ownership & Control:
   1.1. Printed version of the journal - US$260 / €220.
   1.2. Electronic version - US$185 / €150.
2. Corporate Board: role, duties & composition:
   2.1. Printed version of the journal - US$240 / €190.
   2.2. Electronic version - US$180 / €145.
3. Risk Governance & Control journal: financial markets & institutions:
   3.1. Printed version of the journal - US$240 / €190.
   3.2. Electronic version - US$180 / €145.
4. Journal of Governance and Regulation:
   3.1. Printed version of the journal - US$240 / €190.
   3.2. Electronic version - US$180 / €145.

For institutional subscribers:

1. Journal of Corporate Ownership & Control:
   1.1. Printed version of the journal and electronic subscription - US$1240 / €890.
   1.2. Printed version of the journal - US$1080 / €790.
   1.3. Electronic version - US$730 / €650.
2. Corporate Board: role, duties & composition:
   2.1. Printed version of the journal and electronic subscription - US$820 / €760.
   2.2. Printed version of the journal - US$960 / €720.
   2.3. Electronic version - US$650 / €590.
3. Risk Governance & Control journal: financial markets & institutions:
   3.1. Printed version of the journal and electronic subscription - US$920 / €790.
   3.2. Printed version of the journal - US$860 / €620.
   3.3. Electronic version - US$550 / €490.
3. Journal of Governance and Regulation:
   3.1. Printed version of the journal and electronic subscription - US$920 / €790.
   3.2. Printed version of the journal - US$860 / €620.
   3.3. Electronic version - US$550 / €490.

Underline one of the payment methods you prefer, and write amount to pay (if you prefer, you can pay by one cheque/bank transfer to subscribe to both journals):
1. I enclose a cheque for US$ / €_____;
2. Send me an invoice for US$ / €_____ (indicate currency correctly).
Write your contact details here:

Name_____
Position_____
Institution_____
Address_____
E-mail_____Tel_____Fax_____

Please, send this form (with a cheque if you prefer to pay by cheque) at:

Dr. Alexander Kostyuk
Publishing house "Virtus Interpress"
Postal Box 136
Sumy
40014
Ukraine
Ask for invoice by fax at +380-542-698125 (if you want to pay by bank transfer).
Enter our web-site to download the subscription form at www.virtusinterpress.org.