10.14751/SZIE.2017.056

Szent István University

PhD School of Management and Business Administration
Gödöllő

Doctoral (PhD) Dissertation

Data Science to Improve Bank Controlling Methods

Written by:
Péter Kalmár

Supervisor:
Dr. Zoltán Zéman
associate professor

Gödöllő
2017

10.14751/SZIE.2017.056

The PhD School

name: **PhD School of Management and Business Studies**

scientific field: **Management and Business Sciences**

head:
Dr. József Lehota
professor
Doctor of the Hungarian Academy of Sciences (MTA), economic science
Szent István University
Faculty of Economics and Social Sciences
Institute of Business Studies

supervisor:
Dr. Zoltán Zéman
associate professor
PhD, economic science
Szent István University
Faculty of Economics and Social Sciences
Institute of Business Studies

..... **.....**

Approved by the Head of School **Approved by Supervisor**

Table of Contents

"Research is to see what everybody else has seen,
and to think what nobody else has thought."

Albert Szent-Györgyi (1893 – 1986)

1. INTRODUCTION

1.1 THE RELEVANCE OF THE TOPIC

The bank sector has a vital role in the modern economy. It is a sensitive monitor, engine and control factor of the economic system. The banking industry can directly influence the production, consumption, investments and indirectly other important areas such as the innovation, education, etc. through monetary solutions, wide range of banking products and financial services for the different members of the market. For example, in the second quarter of 2014 the financial services[1] provided 9.2%[2] of the gross value added (GVA) output in the United Kingdom parallel with this in Hungary the sector had 4.3%[3] contribution towards to the gross domestic product (GDP) in 2012. Without doubt the recent changes in the Hungarian bank system can highlight the importance and influence of the financial system in the economic situation of a country. During the recent (2008-2012) financial crisis due to the increased risk levels (country risk, operational risk, etc.) the stock of deposits funded by Monetary Financial Institutions (MFI) of the European Monetary Union (EMU) decreased by 14.8%[4] between January 2008 and December 2012 *(the figure is 46.9%[4] between January 2009 and December 2012).* It directly impacted the stock of loans to non-financial corporations and businesses which have dropped by a significant 7.4%[4] for the same period *(and 19.3%[4] for the period between January 2009 and December 2012)* which (with other factors) led to less new investments and developments *(10.7%[5] reduction in national investment between 2009 and 2012).* As this example demonstrates the bank system is an organic part of the modern economy.

Following the recent financial crisis, the focus is on re-regulations, the measurements of risks, efficiency and profitability as well as on the value adding processes for the investors and the customers. The external regulatory system has been changed significantly over the last 4 years. [TÓTH, 2013] The new primarily risk focused Basel Accord (Basel III.) was introduced and new regulatory bodies have been created. For example, in the United Kingdom the previous Financial Services Authority was replaced by three new authorities (Financial Policy Committee, Prudential Regulation Authority and the Financial Conduct Authority) for a more thorough regulation of the banking industry. At the same time banks also made several internal changes; on the one hand banks reviewed - the already indispensable part of their business - their risk management which *"designates the entire set of risk management processes and models allowing banks to implement risk-based policies and practices."* [BESSIS, 2002, p. 2]. These allow to measure, monitor and control risks. On the other hand, there is a transformation in the system as a result of the recent financial crisis. The focus has shifted from the pure profit- and product-oriented thinking towards to the effective, sustainable and profitable management control where besides the basic services the organisations are

[1] Including insurance and pension funding and activities auxiliary to financial services
[2] Own calculation, source: ONS (http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-319470); last accessed on 14/08/2014
[3] Own calculation, source: KSH (https://www.ksh.hu/docs/hun/xstadat/xstadat_eves/i_qpt002c.html), last accessed on 14/08/2014)
[4] Own calculation, source: MNB (http://english.mnb.hu/Statisztika/data-and-information/mnben_statisztikai_idosorok , "X. Monetary And Other Balance Sheet Statistics" section, "Monetary Statistics" file), last accessed on 25/08/2014
[5] Own calculation the figure includes all national investments, source: KSH (http://www.ksh.hu/docs/hun/xstadat/xstadat_hosszu/h_qb001.html), last accessed on 26/08/2014

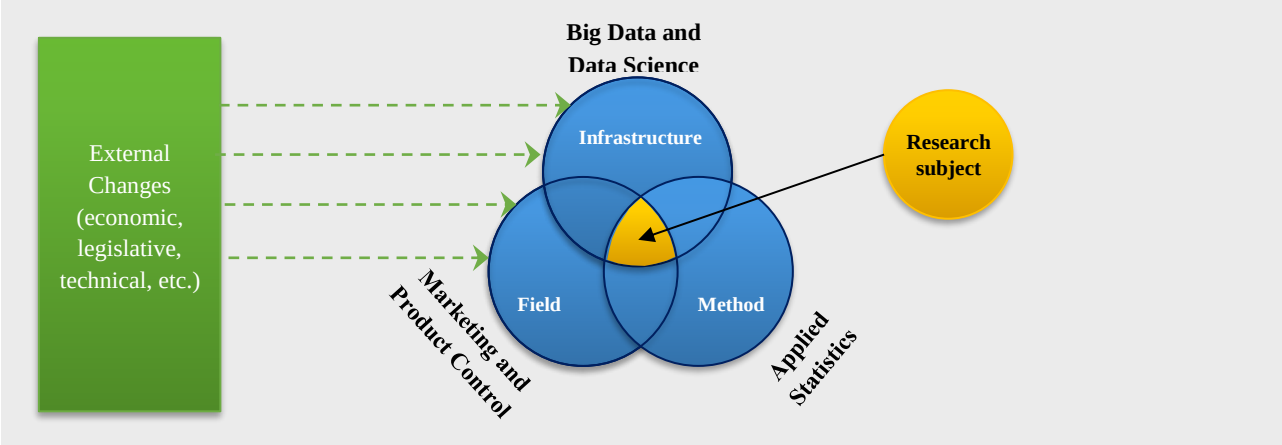
aiming to add value for the investors and the customers. Retail companies like Tesco, Marks & Spencer or Sainsbury's have already recognised this change in the importance of customer loyalty and launched their financial divisions forcing the entire financial industry to change the primarily profit and fund focused mid-term strategies to long term sustainable strategy that is based on quality of customer relationship, satisfaction and profit targets at the same time. In order to achieve this, banks – among other structural and cultural changes - have to develop an effective planning and controlling system that is adapted to the new challenges and focusing on financials, customer and market at the same time rather than on financials only therefore it is key for Banks to use controlling and marketing tools as well as the latest developments of statistical and data sciences together.

1.1.1 Empirical Observations

In the past decade, the author has been in various analytical roles, ranging from credit risk analysis through product and customer behaviour analysis to financial and business analysis at several financial institutions in Hungary and the United Kingdom. During these years, author developed a comprehensive knowledge of the analytical and reporting methods and practices applied by the different analytical functions at banks. Although this experience may not be representative of the industry average, the combination of methods taught at the public accountancy qualification course and the accumulated experience gained as a financial control manager highlighted a potential opportunity to improve existing controlling methods in the planning and evaluation fields. Author found at the financial institutions he worked for that a comprehensive statistical and mathematical knowledge, expertise and daily practice existed however it has mainly concentrated around risk, fraud and customer modelling and was little or not present in other, primarily financial and accountancy related areas. Furthermore, the central management control areas managing the control cycle (plan, do, check, act) are often focusing only on the accountancy side of planning and evaluation although econometric and statistical methods could provide more accuracy and insight into issues which require a complex, multidisciplinary view. Even the widely applied "driver-based" planning is often too static and in some cases based on outdated assumptions therefore require continuous re-planning. Moreover, these planning systems do not assess the interrelations of the different factors but target specific issues (sales, customer loyalty, etc.) separately or without the adequate level of interconnections. The constant need for updating drivers and plans as well as the cost of provisional funds (funds that are effectively not reinvested into the core activity until they released) decreasing the efficiency and essentially costing money and other valuable resources for the organisation. Therefore, any improvement in planning or any insight through the evaluation (or check) phase that can help the business to focus the resource more efficiently is beneficial to the professional and scientific communities at the same time.

This experience provided the required motivation for author to: a.) understand how the planning and evaluation phases of management control can be improved with the use of statistical methods; b.) understand the required steps for the practical development of improved methods; c.) understand if there are any current obstacles and finally d.) understand how to manage or break down these issues. Based on initial consideration author believes that three key components required for a successful implementation of an advanced method: 1.) fast, efficient (and cheap) platform and data processing capabilities; 2.) elaborated and proven method for planning and evaluation and 3.) cultural shift towards to the adaptation of such methods.

In terms of the first point author found that, during this decade the rapid development of data processing technologies, such as "big data" [6] solutions can support not only fraud, customer relationship or digital services but everyday financial and management control processes, planning and evaluation too potentially more efficiently and less expensively than traditional methods. The new era of data science enables professionals with detailed statistical (specifically machine learning), computer coding knowledge and expertise from their respective area (such as financial planning or credit risk, etc.) to develop new, improved methods to support business decisions. [CURREY, 2016] In terms of the second point, as the data processing infrastructure already exist a research should be conducted to understand: a.) whether the different goals and targets of financial institutions can be modelled with the system of costs, expenses and other factors; b.) if such models can provide new insight into the complex interrelations of factors and ultimately change the focus and strategy (therefore rationalise the efforts and resources) if required and; c.) whether the sequence of required methods and steps can be captured on process maps for financial and control managers and analysts. Although such improvement can potentially be applied in several parts of cost planning and forecasting environment the current research is focusing on the improvement of marketing cost planning. Author is looking to answer questions related to: a.) the key factors that determine the success of a bank marketing campaign aiming the recruitment of new customers or the increment of market share; b.) the relationship between marketing spend and campaign aims to predict the cost of certain campaigns or to evaluate the success of past campaigns. The research is focusing on bank marketing campaign costs due to the complex nature of the topic, the importance of customer focus and the current emergence of a new marketing and product centred management controlling sub-function. Figure 1 shows the key dimensions of the research subject and opportunity.



Source: Compiled by author

Figure 1: The key dimensions of the research topic

As we can see the role of financial industry is vital for the economy. Following the recent financial crisis, the emphasis is on efficiency, re-regulations and on new long term strategies that create value not only for investors but for customers and other stakeholder too. This environment combined with the very recent developments of data sciences creates the opportunity to improve existing controlling methods in the areas of planning and evaluations. In current doctoral dissertation Author is dedicated to take this opportunity and respectfully provide both the academic and the professional communities with the option to use an improved planning and evaluation methods.

[6] *"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization."* [LANEY, 2001]

1.2 THE SCOPE OF THE DISSERTATION

Current dissertation is synthetizing the results of author's research activity from the past 5 years. The research is centred around the following three, interrelated key topics:

1. *The new bank marketing-controlling function in practice*: Author has been reviewing the literature of management accountancy to understand the function, impact and applications of the control concept in the field of marketing management. The dissertation is primarily focusing on the controlling aspects of this fusion and approaches marketing from a statistical and financial analysis angle therefore the full review of marketing management and its functions is out of scope of current research.

2. *The application of statistical methodologies in traditionally accounting driven management control areas, such as financial planning and actual versus budget analysis*: Author would like to understand the current theoretical and practical planning and evaluation methods used by the marketing management control function. Based on the assumption that these procedures primarily focus on non-statistical methods[7] author is assessing the option to apply simple statistical methods such as multivariate methods (e.g.: regression analysis, factor analysis etc.) to offer an alternative procedure that improves existing planning accuracy and provides more insight (compared to indexing for example) to decision makers.
   Current research is focusing on an alternative procedure that supports marketing cost planning and the post campaign actual versus budget analysis for primarily current account banking products but also considering credit card and savings products. There are two campaign aims considered for model building and post campaign analysis: to increase number of customers by the targeted volume; and to reach targeted market campaign share. It is important to highlight that the proposed procedure is generalised so it can be used for other campaign aims as well, however other goals do not form part of the current research scope.
   The simplicity and efficiency of the new procedure are very important aspects as the improvement is intended to be used primarily by financial controllers and not by statisticians. Therefore, Author is assessing if the simple method of multivariate linear regression is the appropriate approach. Furthermore, if the process can be captured by simple process mapping for re-usability. The application of non-linear regression modelling methods is out of scope of the current dissertation as well as other advanced, computer based forecasting methods such as neural networks and decision trees.

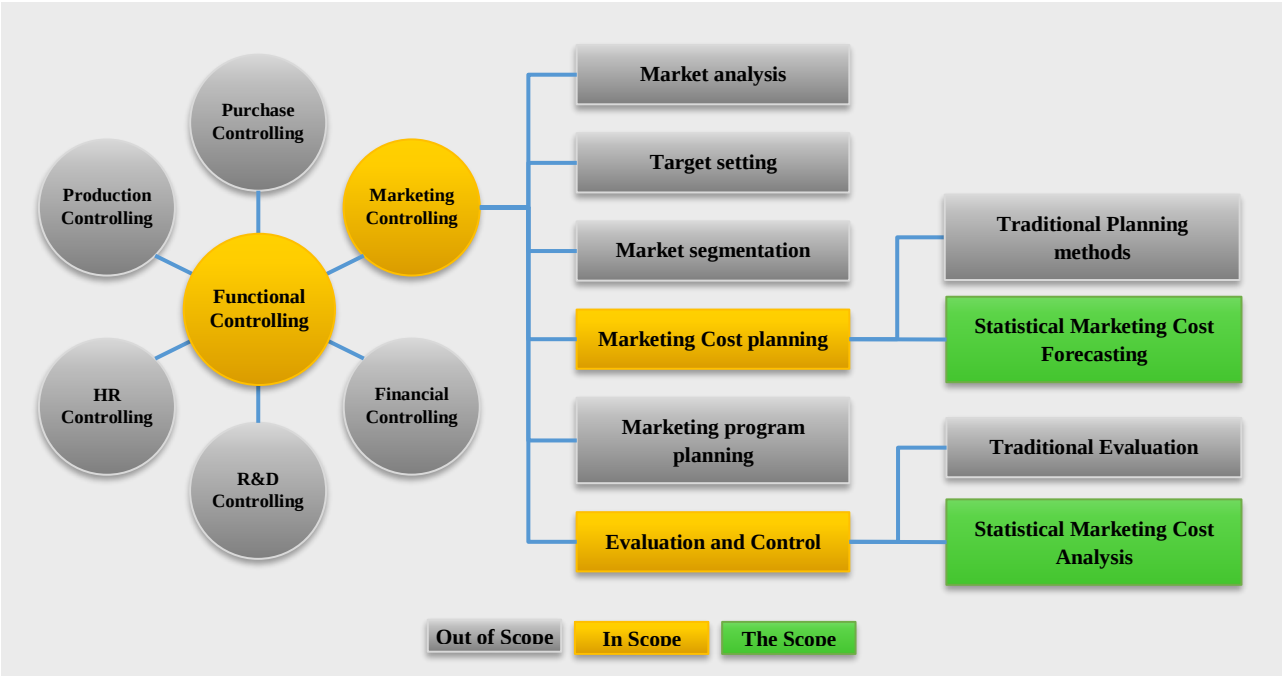
3. *The practical strategic and operative benefits of (or lessons learnt based on) the fusion of data science, applied statistics and marketing management control:* Applying statistical methods can extend the analytical capabilities of the financial controller as a wider range factors, multiple dimensions of the internal and external environment of the organisation can be analysed simultaneously due to the increased data capturing, storing and processing capabilities provided by data science. This enables financial controllers and managers to make more informed decisions and to gain insight into the complex interrelations of different factors. Author in the current study would like to establish the key contributing factors and their weights that influence the customer recruitment volumes and campaign market share. For this author is using an initial

[7] For planning for example accountancy methods such as zero based planning, for evaluation for example indexing.

range of potential influencer (or predictor) variables collected from an experienced focus group of financial professionals working in the bank sector of the United Kingdom. This set of variables are complemented with variables from the literature however author is only considering 200 initial variables as the scope of current study. In an earlier research [KALMÁR et al., 2015] author identified the key factors that determine the re-purchase propensity of savings product. This information, similarly to the key contributing factors of the campaign success and their quantified influence, enable managers to identify key areas of focus and support a more efficient resource management ptactice.

Further to the above-mentioned limitations, the model presented in dissertation was completed based on simulation data desensitised from actual industry data collected from one of the largest[8] financial institutions in the United Kingdom. Given the sensitivity of this information author's only option to produce a publicly available, realistic and practical method built and tested on actual industry data is if the data was desensitised. Testing the model in other countries, primarily in Hungary, is out of scope of the current study.

Figure 2 details the identified knowledge gap in the controlling practice. The novelty of current research is the application of statistics in a traditionally accountancy dominated functional controlling area - the marketing management control area - to improve the accuracy and efficiency of marketing cost forecasting, marketing campaign evaluation and analysis with a proposed new procedure based on existing but simplified statistical methods. The yellow boxes in Figure 2 are highlighting the areas covered and impacted by the research, shown from top level to the bottom granularity (from left to right on the chart below). The green boxes represent the new methods. Not only the procedures but the outcome of the applied methodology will contribute to scientific and business knowledge.



Source: Compiled by author based on [GAÁL, 2007., pp. 2.] and [VÉRY, 2009., pp. 7.]

Figure 2: The identified knowledge gap and the scope of research

[8] The financial institution is in the top 10 financial providers of the United Kingdom based on the total asset and has over 10% share of the current account market at the end of December 2016.

1.3 AIMS AND OBJECTIVES

The aims and objectives of the dissertation based on the previously outlined scope are the following:

I. Aims and objectives related to the literature review

1. *The review of existing "Controlling, Management Control" literature. Definition of controlling concept based on the Hungarian and Western literature.* Objectives: *a.)* define controlling in today's practice and clarify the naming convention in different languages; *b.)* explore the evolution of controlling concept by identifying key milestones led to current practices; *c.)* review the role of management control in the modern economy, specifically its interactions with the legislative environment.

2. *Definition of the system, aims and functions of Management Control. Examination of the Financial Institution Controlling specifications.* Objectives: *a.)* outline the functions and objectives of controlling highlighting the difference between monitoring and controlling systems; *b.)* describe the general model of a controlling system; *c.)* define the concept of Financial Institution Controlling; *d.)* review the specifications of the controlling system.

3. *The review of existing Marketing Management Control approaches.* Objectives: a.) review the relevant marketing management theories, concept and functions; *b.)* outline the interrelation between marketing management and the controlling concept; *c.)* define Marketing Controlling and highlight functional specifications; *d.)* outline the system, functions and objectives.

4. *Description of the functions and processes of Strategic and Operative Planning and Supervision in Management Control.* Objectives: *a.)* evaluate current methodologies and processes; *c.)* assess the marketing and product sales related to planning and evaluation processes, especially to applied key performance indicators (KPIs); *d.)* examine sales planning, monitoring and evaluation; *e.)* examine cost (specifically sales related campaign cost) planning, monitoring and evaluation in the literature.

5. *The review of the latest developments of data sciences in the financial industry.* Objectives: a) define the concept of big data and data science; b) detail relevance to the topic; c) list existing applications in the financial industry; d) describe potential applications of data science (e.g. campaign cost prediction) in relation to marketing cost predictions and marketing controlling.

II. Aims and objectives related to research

1. *The methodology review of existing Management Control cost planning (especially marketing budgeting) and evaluation procedures.* Objectives: *a.)* identify key authors from Western, German and Hungarian literature based on the above literature review; *b.)* collect and review the described cost planning and evaluation methods of the selected authors *c.)* categorise their methods based on the level of mathematics and statistics involved.

2. *Examination of existing cost planning and evaluation procedures to see if they can be improved with data science and statistics.* Objectives: *a.)* establish the maturity of the combined literature of data science and management control; *b.)* test if a mathematical/statistical function exists between key bank marketing goals and campaign costs with other contributing factors; *c.)* assess the potential application of the function in campaign cost prediction and evaluation.

3. *The establishment of the end-to-end*[9] *procedure (unique sequence of statistical methods and decisions) to support financial analysts and controllers to conduct marketing campaign cost prediction and evaluation.* Objectives: a.) review literature of linear modelling and statistics; b.) process the literature related to cost predictive procedures; c.) process literature of other multivariate methods; d.) establish main stages of (campaign) cost prediction modelling; e) design simple process flow to support financial analysts and controllers to conduct marketing campaign cost prediction and evaluation; f.) detail statistical terms and required knowledge.

4. *Identification of the factors that determine the success of banking or savings product sales campaign. Analysis of their contribution to the success based on practical data.* Objectives: *a.)* collect all available related variables using data mining techniques; *b.)* identify target variables (e.g. maximum sales per campaign, maximum market share per campaign); *c.)* identify the key factors (using multivariate methods) that determine target variables; *d.)* identify key variables based on the factors; *e.)* analyse the relationship between variables and determine their contribution to the result (using regression analysis); *e.)* explain the results scientifically.

5. *Conduction of comparative analysis between the existing campaign cost evaluation/planning process and the improved cost evaluation planning approach.* Objectives: *a.)* set KPIs to measure the efficiency of the existing campaign cost evaluation/planning process; *b.)* measure the same key indicators in case of both processes; *c.)* analyse and explain results.

Table 1: Summary of research aims and strategy

Aim	Hypothesis	Research Strategy	Applied Methodology	Results
Review campaign cost evaluation and planning processes.	Cost planning and evaluation methods are dominantly non-statistical, accountancy based.	Secondary research	Literature review	Chapter 4.1
Assess if enablers of improvement are available in literature	The literature of management control harnessing big data technology and data science is still developing.	Secondary research	Literature review	Chapter 4.1
Examine if the existing planning and evaluation process can be improved with data science and statistics.	A mathematical function exists between key bank marketing goals and campaign costs with other factors.	Primary and Secondary research	Data collection, statistical experiment	Chapter 4.2
Establish the end-to-end process to support analysts and controllers to conduct marketing campaign cost prediction and evaluation	The mathematical function can be applied for cost prediction and evaluation. The analytical process can be captured on process diagrams.	Secondary research	Literature review, Process mapping	Chapter 3.3
Identify the factors and their weights that determine the success of banking or savings product sales campaign	There is a linear relationship between campaign spend marketing goals. The success factors and their weights can be modelled.	Primary	Modelling, Multivariate methods	Chapter 4.3
Evaluate the new process compared to existing planning methods	The application of predictive cost model is increasing planning accuracy.	Primary	Descriptive statistics	Chapter 4.4

Source: Compiled by author

Table 1 summarises the aims of current dissertation. As detailed above current dissertation not only aims to assess the feasibility of statistical prediction of marketing campaign costs but also to deliver to both the scientific and the business communities insight into the success factors of bank marketing campaigns and to provide the blueprints of the methodology for practical applications.

[9] *"End to end is a term used in many business arenas referring to the beginning and end points of a method or service […] it ensures the comprehensive completion of work and is typically managed within a specified timeframe."* For more details please visit: http://www.investopedia.com/terms/e/end-to-end.asp (Last accessed: 29/10/2016)

"Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so we may fear less."

Marie Curie (1867 – 1934)

2. LITERATURE REVIEW

The word "kontrolling" used in the Hungarian terminology is coming from the German word "controlling" which is adopted from the English "to control" verb for "control" and "regulate". The original name of the concept may have come from the French word "comptroller" (invoice checker, Controller) which is a fusion of the words "compte" (account) or "compter" (count, counter) and "controleur" (inspector, auditor), "contrôler" or "contrer" (check). The English names of the concept used today are "Managerial Accounting", "Financial Controlling" but also referred to as "Management Control" [KÜPPER, 2007, p. 750].

Although the origin of cost recording goes back to the ages of royal treasurers of the 13th century England and France [HÁGEN, 2008], the early Management Control approach was developed during the Industrial Revolution in the United States of America two centuries ago. The newly formed large production factories and the fast-growing transportation technologies of the 19th century required the continuous invention and development of ratios and measurements. [CHANDLER, 1977] The real breakthrough came in the 20th century when the early Management Control approach became a business discipline. Even though the foundations of the modern management controlling were laid down by the du Pont cousins, who consolidated all of the latest methods into the famous DuPont ROI management account model [KAPLAN-JOHNSON, 1987], the contributions of Henry Ford and Alfred Sloan to the practical and theoretical concept development are often considered as revolutionary. [HAMMER-CHAMPY, 2001]

In the second half of the century companies worldwide started to adapt the management control practice resulting differences in the development of the concept due to the diverse focal points of the local control practices (which is ultimately was led by the culture of the nation). For example, while in German-speaking countries the control practice mainly focused on the accountancy and business administration aspects, the Western-type control concept was built around management and management support. [SZÓKA, 2007] The management control, or "kontrolling" as often referred to in the German literature, in the last two centuries became a business discipline and a scientific field due to the great contribution of authors like F.W. Taylor, Henry Ford, Alfred Sloan, Robert Kaplan, David Norton, Robert Anthony, Hans-Ulrich Küpper and Peter Horvath amongst others.

2.1 THE ROLE OF MANAGEMENT CONTROL IN ECONOMICS

The concept of Management Control has evolved since its first appearance due to the different economic challenges of different times. Early studies described Management Control as a formal system with the financial and accounting information systems in the centre. The main focus of early adopters was on cost accounts and budgets. More recent studies have stressed the importance of human relationships, leadership, motivation and the organisation's culture as less formalised aspects of control systems. Management Control today is a modern control system with social connotations and is open to the influences of the organisation members and its environment. [CARENYS, 2012]

2.1.1 The History of the Management Accounting and Control

"Watch the costs and the profits will take care of themselves." (Andrew Carnegie)
[In: CHANDLER, 1977, p.268]

The history of Management Accounting and Control can be traced as far back as the end of the 13th century when Royal treasurers (called "pursebearers") in England and France recorded direct and standard costs of the Kingdoms. [HÁGEN, 2008]

The early Management Accounting approach we know today was developed during the Industrial Revolution in the United States of America two centuries ago as a result of business owners exploring different ways to improve their comparative advantages by capturing and using information about their organisations. The newly formed large production factories of the 19th century required effective planning which was only possible with a management accounting function that was capable of capturing and providing the required information to plan, control and evaluate the new processes. Later in the century new technologies enabled railroad companies to expand their routes nationwide. In order to maintain the required profit levels and to evaluate the cost of transportation executive accounting managers started to develop long-distance cost performance ratios and measurements. [CHANDLER, 1977]

The evolution of management accounting continued at the beginning of the 20th century. With the appearance of the wholesalers and retailers (for example R.H. Macy & Company Inc. in New York or the Sears, Roebuck & Company in Chicago) the scope of management accountancy expanded. Managers started to focus on asset and inventory management measures amongst other existing ratios and metrics. Understanding the processes and the flow of raw materials became an important element of cost ratio calculations. For example, the engineer Frederick Taylor assessed the existing workflows in order to simplify and rationalise processes. With the help of company accountants, he was able to calculate accurately the hourly wage cost and cost of specific raw materials needed for production. [KAPLAN-ATKINSON, 2003] In fact a number of contributions to the field of management accounting came from disciplines such as economics and engineering. Another contribution for example was of Henry Hess, a mechanical engineer who developed the first breakeven chart. [PARKER, 1969] The first comprehensive management accounting methodology was elaborated by the du Pont cousins (in 1903) who consolidated all the latest methods into the famous DuPont ROI management account model.[10] [CHANDLER, 1977]; [KAPLAN-JOHNSON, 1987]

In the first half of the 20th century the first official, professional organisation to be established in the field of Management Accounting and Control was the Controllers Institute of America in 1931.[11] In 1940 the same institute issued the first general list of controllership tasks[12], which primarily covered the areas of accounting, tax administration, planning for control, reporting and interpretation, protection of assets, financial transactions and government reporting. The development of the classical management accounting concept reached its peak by the work organisation and industrial engineering principles of Henry Ford, and Alfred Sloan. The starting point for both of them was Adam Smith's theory. In the Wealth of Nations, Smith, a radical thinker and forbear of business

[10] The model was later extended by General Motors in 1920.
[11] The expansion of responsibilities of financial executives into policy-making areas led the organisation to change its name to Financial Executives Institute in 1962. Read more: www.financialexecutives.org
[12] The list was reviewed and re-published in 1962.

consultants, explained what he called the principle of the division of labour. During his visit to a pin factory Smith observes that *"some number of specialised workers, each performing a single step in the manufacturing of a pin could make far more pins in a day than the same number of generalists, each engaged in making whole pins. [...] the division of labour increased the productivity of pin makers by a factor of hundreds."* [HAMMER-CHAMPY, 2001, p.14] While Ford applied the logic for the work organisation, Sloan adapted this principle to the management. His view was that managers of the organisation should not be required to have specific engineering or production management skills, mainly the management of functional areas should be assigned to specialists. [BODA et al., 2001] Table 2 summarises the key focus, developments and authors of the described periods of the management control evolution in the USA.

Table 2: The Western history of the management accounting and control

	Early Accounting Period	Early Management Accounting Period	Classical Management Accounting Period	1st Management Control Revolution	2nd Management Control Revolution	Post-modern Period
Dates	Before 1700	1700-1900	1900-1950	1950-1980	1980-1999	After 2000
Focus	cost recording	capture and provide cost and process information	accounting, tax administration, planning for control, reporting	reporting and decision-making support	support information need of managers	data science, digitalisation, culture based control
Achievements	split of direct and standard costs	cost-performance ratios	management accounting became an official profession	management accountancy became part of academic curriculum	division of management control and accounting	shift from mechanistic view to complex set of organisational interdependencies
Pioneers	-	du Pont cousins	Taylor, Sloan, Ford	Robert Anthony	Kaplan, Johnson	various

Source: Compiled by author

According to Birnberg [1999] in the 1950s the "classical period" of management accounting evolution[13] ended and the "modern" management accounting and control period started with the first management control revolution (1950-1980). In the United States of America the field of management accountancy became part of academic curriculum at Harvard University and Massachusetts Institute of Technology. [MAHER, 2000] Economist and accountants started to develop new techniques and decision-making tools to solve the traditional problems of improved profit and efficiency. The first influential textbooks on management accounting were written by Robert Anthony in 1956[14] and Charles T. Horngren in 1962[15]. These books were the first to primarily focus on the decision-making and management aspects of accounting. During the second management control revolution (1980-1999) the rapid development in the field of mathematical modelling influenced more and more researchers to target Management accounting and control issues. The new techniques and solutions highlighted that management accounting curriculum on its own was no longer able to support the new challenges managers were facing. [BIRNBERG, 1999] Leading researchers wanted to "e*xtend management accounting into nonfinancial areas; a better understanding of contemporary problems and the information needs of managers"* and also called for "*innovative practice"*. [KAPLAN-JOHNSON, 1987, p.47] The division between management

[13] A period between 1700 and 1950s.
[14] Management Accounting published by Richard D. Irwin, Inc.
[15] Cost Accounting: A Managerial Emphasis published by Prentice-Hall

accounting and management control was more pronounced towards to the end of the century. The period since 2000 in the US became known as the "post-modern period" and it is best characterised by the shift from the mechanistic view of organisations to acknowledging a complex set of organisational interdependencies. [BIRNBERG, 2000, p.717]

The European[16] history of Management Accounting and Control is slightly different to the mainstream evolution. Although the early "controlling" concept can be traced back to the appearance of first German large-corporations and factories, the main expansion of the concept started with the rebuilding of Europe after the Second World War. By the end of the 1950's "controlling" became an independent function mainly focusing on the accounting elements with the "Controlling of the Operations" in the centre. [SCHMALENBACH, 1963] The development of the controlling literature in the German-speaking countries started only later in the early 1970's. The key focus of the management accounting and control function was to provide the required information for decision making as well as to coordinate this information flow. The first official, professional organisation in Europe was the Controlling Akademie established by Albrecht Deyhle and his company in 1971[17]. Due to the rapid development by the 1980's the German-type cost calculating systems were outdated and got criticised by researchers as such systems were lacking the consideration of value adding processes; the identification of the real source of costs; and the lifecycle approach. In order to adapt to the new challenges German controlling researchers and professionals developed new methodologies. The new approach was called Process-Cost Calculation, which applied similar principles to the Western ABC (Activity-Based-Costing) method. Also in the same decade, a new, management oriented approach appeared in the German control practice. In this context, controlling became the tool of the coordination function of management. [LÁZÁR, 2002, p.41]

Other parts of Europe have shown a different paste and path of development. Although the "controlling" concept appeared in Central European countries like Hungary in the early 1960's, the real development of such systems was delayed until the end of Communism (1989). As the artificial organisation hierarchy within the large post-communist corporations started to ease, the different organisational units became more independent which required a sophisticated coordination mechanism. Naturally managers started to look for and apply new practices resulting an innovation "push" effect. After the gradual transition into market economy, the accumulated knowledge and experience of Western (mainly German) companies and investors reached the region and resulted a fast and widespread deployment of controlling practices, an innovation "pull" effect. During the decade of transition the new Hungarian specific branch of Management Accounting and Control was formed. [CHIKÁN, 1997 p. 26.] Despite the efforts of the professional community, unfortunately management control in Hungary is often adopted mechanically and used in isolation. Besides, Hungarian companies often limit management control practices to the coordination of accounting related questions and do not consider it on a strategic level.

2.1.2 The Management Control Concept

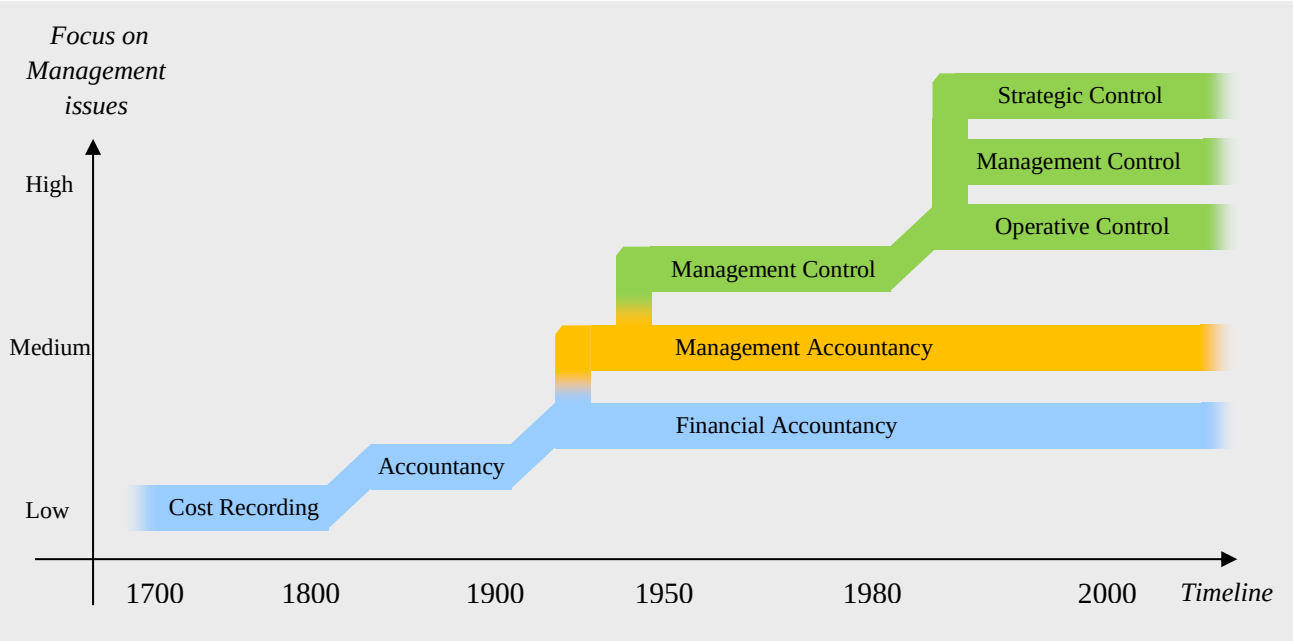
As we can see from the previous chapter, the history of Management Control shows several differences due to various interpretations and needs of different cultures. Since the 1950's there have been two (Management Control) mainstreams existed in management sciences: the original Western

[16] Mainly in Continental Europe
[17] Read more: http://www.controllerakademie.de/english/english.html

(or "English") type, which started to develop in England and the USA during the industrial revolution; and the Continental European (or "German"), which has existed since the appearance of large production factories but mainly developed after the Second World War. Both types share common principles. However, there are key differences mainly driven by the different application and the approach of management control. Despite of these differences in the age of data science and global knowledge transfer the "*development moves towards combination of the two types*" [HANYECZ, 2006, p.7] where both trends are interacting.

The need for an overall control concept emerged when organisations started to focus on certain end goals or objectives. According to Chester Barnard [1938, p.8.] "*the organisation is a system of consciously coordinated activities [...] which are conducted with certain explicit purposes*". In order to check whether these explicit purposes or objectives are achieved some kind of evaluation system is required. A system measuring these achievements will not only impact the behavioural patterns of the top management but it is also vital for the workforce at the lower levels of the organisation as tasks at lower level may not seem to directly serve the overall objectives. Therefore, the evaluation system can help with the motivation of the person in charge of the relevant task. [ROSANAS, 1994] Furthermore, the information of such systems helps managers to develop and maintain viable behavioural patterns and to control performance. [OTLEY, 1999]

The first western organisation theorist to introduce the term of 'management control' was Robert N. Anthony [1965]. He interpreted the concept as a collection of processes in line with the corporate goals and objectives that ensure the effective and efficient acquisition and use of resources. Before Robert Antony's definition - at the early stages of the concept formation - management control and management accounting were considered as synonymous terms. [OTLEY, 1999] The main aim of such systems was to produce periodic and non-periodic reports and special information for internal and external stakeholders to support operative, strategic or tactical decisions. [HORNGREN and FOSTER, 1991] As Figure 3 demonstrates the real distinction started only in the late 1960s when Western researchers realised that a flexible information and coordination system - primarily focusing on issues important for managers - was required.



Source: Compiled by author

Figure 3: The evolution of Management Control Concept by the level of focus on management

Although Management Accountancy was already focusing on the planning and control and was future orientated, the key source of information was still the primary accounting system. Instead of the mechanic accounting rules and principles Management Control started to offer a wide range of innovative mathematical solutions and broader integration of information sources. The English literature also differentiates between 3 levels of the management control: strategic control; management control; and operative control. Table 3 details the specifications of the tree levels which co-exist in the organisations.

Table 3: The three levels of management control

	Strategic Control	Management Control	Operative Control
Specification of Issues	Complex, not defined / unstructured, multiple alternatives, causal links	Reoccurring tasks / examples, limited number of alternatives, partial automation	Predefined rules, application of mathematical models
Time horizon	Depending on the industry it can be decades	The next couple of years, the focus is on the forward-looking approach	Near future
Specifications of the Control Process	Less formal analysis, importance of management decision, there is enough time for thorough analysis, unregulated	More formal analysis, there are deadlines, less iteration	Formal analysis, reoccurring, follow the rules
Evaluation / Analysis	Subjective and complex, evaluation is only possible on long term	Less complex, regular (min.) annual re-evaluation	Based on clearly defined set of measures, instant evaluation
Focus	Assessed unit or long term plan (program), less hierarchical	Program and responsibility centres, hierarchical	Unique transactions, not hierarchical
Deadlines	Speed is generally not important	Speed is more important than accuracy	Immediate

Source: Robert N. Anthony in Francsovics, 2005., pp. 47.

Strategic Control Concept: The strategic controlling helps the management and executive decision makers to adapt flexibly to changes in the environment while supporting the strategic aims and the efficient operations. [SZÓKA, 2007, p. 98] The strategic management tasks of the senior management are supported by the strategic functional controlling. [KÖRMENDI AND TÓTH, 2006, pp. 73.] The main responsibilities of the strategic control sub-system are: supporting the formation of the strategic aims; disaggregating the strategic goals to operative targets; constant monitoring, analysis and regular reporting; supporting the modifications of the strategy when needed.

Management Control Concept: The Hungarian literature often uses Management Control and Controlling as synonymous words [FRANCSOVICS, 2005] however there is a clear distinction in the Western practice. Management control is the process of managers influencing the members of the organisation in order to achieve the strategic goals of the organisation. [ANTHONY, 1988] Management control sits between strategic and operative control. It provides managers with the required level of information and control to support the execution of long term (3-5 years) strategic aims by delegating execution and control over the short (1 year) and mid-term (1-3 years) strategies. Generally, speed is more important than accuracy as managers are required to make quick decisions on operative objective when the environment is changing or amend annual strategies in order to achieve overall strategic aims of the organisation.

Operative Control Concept: Operative controlling is responsible for the planning, analysis, information distribution and reporting of the operational activity derived from the strategy. From an accountancy point of view "*operative controlling is the management of the events that have*

associated cost and return in the present". [HORVÁTH, 2001, pp. 141.] Operative controlling tends to be predefined, focusing only on the near future and the present. Deadlines are immediate for analysis and evaluation based on clearly defined measures.

2.1.3 The Definition of Management Control

As organisations evolved the complexity of challenges and needs have changed resulting a conceptual development in the field of organisational control. There were numerous interpretations of the early controlling concept, each adding a new element and helping to distinguish the concept of management control from management accounting. As Carenys [2012] highlights most of the Western authors in the field of management control research agree that management control is a process used by managers subjectively in order to influence the performance and behaviour of the people forming the organisation. [TANNEBAUM, 1966]; [COLLINS, 1982]; [FLAMHOLTZ, 1983]; [FISCHER, 1995] The aim of this management process is to achieve the objectives in support of the overall strategy [ANTHONY, 1965]; [COLLINS, 1982] with the effective and efficient use of the organisation's resources [ANTHONY AND DEARDEN, 1980].

More recently in the European, specifically in the German organisation practice Management Control is often referred to as management subsystem, a separate entity within the organisation that embraces the processes of planning, control and the coordination of information management. [HORVÁTH, 1997] Spremann [1992] highlights that while the management is responsible for the results, the control system is liable for the transparency, which can be achieved by a range of information management, decision supporting and coordination functions. Authors focusing on the information management aspect see management control as a tool, an information management system that supports the organisation to achieve the objectives through management based on the analysis of plan versus actual. [SCHWALBE, 1990]; [MANN AND MAYER, 1993]

Similarly, the Hungarian literature often highlights the Management Accounting elements of Management Control. Mainly focusing on the planning and information management with the accounting information system in the centre. Hungarian authors with strong accounting background often define Management Control or "kontrolling" as an internal management subsystem placing efficiency and organisational expenditures into the focal point, [KÖRMENDI AND TÓTH, 2006]; [TÓTH AND ZÉMAN, 2003] supporting the senior management through planning (strategic and operative), variance analysis of the budgeted and actual figures, audit and control. [KÖRMENDI AND TÓTH, 2006]; [BODA AND SZLÁVIK, 2001]; [HORVÁTH, 1997]; [TÓTH AND ZÉMAN, 2003]; [HANYECZ, 2006]

During the extensive research of Management Control literature, author noticed that scientific publications often limit the analytical aspect of the management control to the variance analysis of budgeted and actual figures. The most recent industry practice and the applications of data science suggest that the analytical function of the management control, especially in case of strategic management control, covers a wide range of analytical topics:

- continuous evaluation of opportunities to help the organisation to achieve the strategic aims;
- analysis of more effective or optimal business solutions;
- predictive trend and forecast analysis;
- project analysis (for impact and cost);

- financial (P&L) and organisational cost analysis;
- staff and productivity analysis;
- and behaviour analysis (of customer, supplier, etc.) amongst other types.

Based on the results of the non-representative survey with industry professionals conducted for current research a shift in the industry practice has been identified. A shift from Management Information primarily produced on a reactive basis (e.g.: continuous comparison of budgeted and actual figures) towards Management Information and analysis produced on pro- and reactive bases at the same time (e.g. evaluation of new business opportunities) by utilising the latest developments of data science (e.g.: predictive model and driver based management control, etc.).

Current dissertation applies the following definition of the management control concept: Management Control is a collection of strategic and operative processes that cover planning, business analysis and management reporting as well as the control functions of the management. The aim of Management Control is to help the organisation to achieve its strategic objectives by sourcing and providing the adequate level and quality of required management information; creating transparency for all stakeholders; and supporting the decision-making process of the management.

Author recognises that Management Control could materialise in many forms depending on the size and structure of the organisation. It can be a management practice of senior managers; or even a separate centralised or decentralised management control function as long as the functions and the aim specified above are covered.

2.1.4. The Management Control in the modern economy

The management control concept has continued to evolve in the past decades and adapted to the social changes and the challenges in the economic environment. Several authors [CARNEY, 2012]; [AMAT, 1992]; [FLAMHOLTZ, 1985]; [BERRY et al., 1995]; [NEIMARK AND TINKER, 1986]; [ARGYRIS, 1990]; [CHENHALL, 2003] have assessed the current trends of the management control systems. The following three main trends have been identified [CARNEY, 2012]:

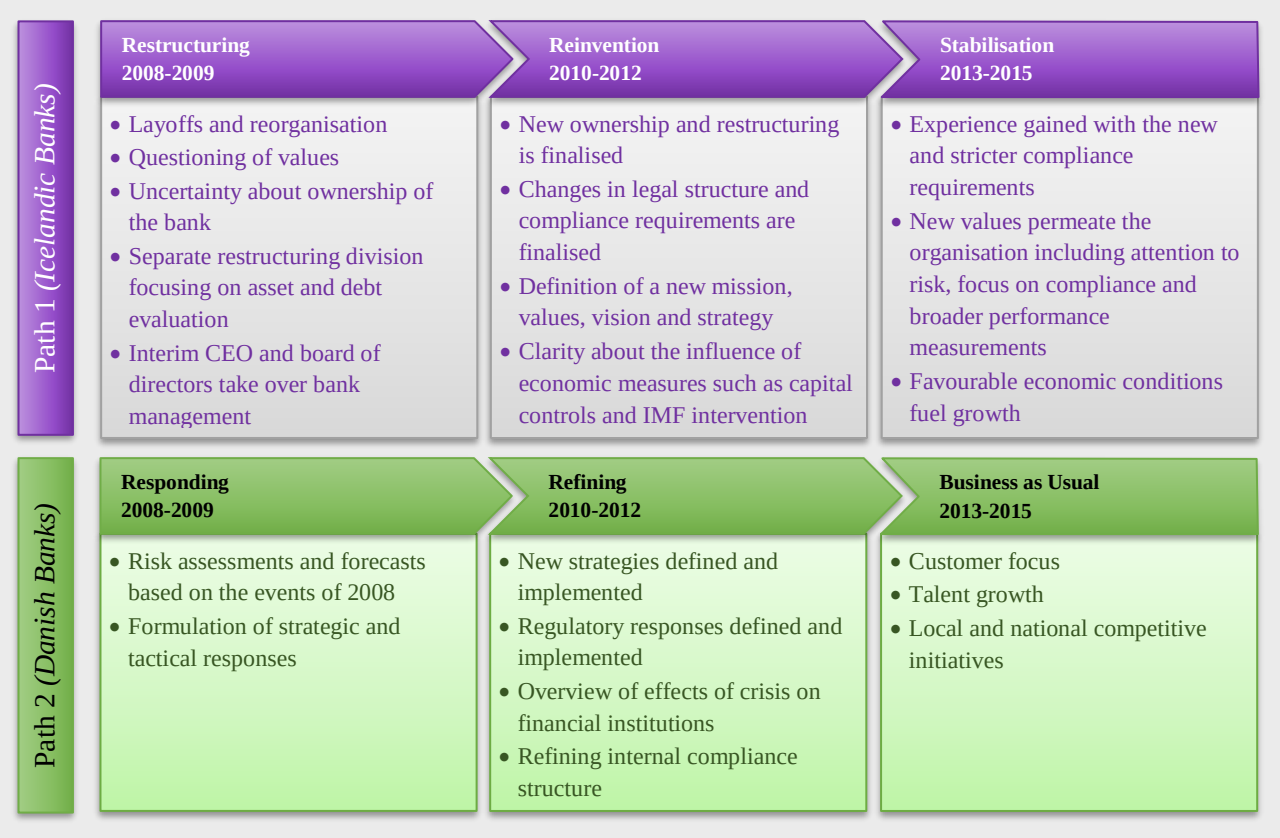
1. **Mechanistic and Formal Control Systems:** such systems are representing the early classical organisational and control approaches [FAYOL, 1949]; [TAYLOR, 1911] where the controlling systems are formal and mainly have technical function in the management activity [CARNEY, 2012]. The organisational objectives behind this type of systems are well-defined, tend to be in relation to profit maximisation and most likely to be planned and controlled by the management. Such organisations tend to be hierarchy driven, behavioural norms and procedures are defined centrally. Most of these systems are based on the principle of *"control by exception"* and the motivation is largely monetary incentive based. [AMAT, 1992] The Mechanistic and Formal Control Systems often fail to take the environment into account or to sufficiently specify its influence on the control [NEIMARK AND TINKER, 1986]. Besides, they only work optimally if the activities are specific and repetitive in a stable environment [AMAT, 1992]. Due to their hierarchical nature, they can prompt demoralising effect on employees at the lower end of the structure [MORGAN, 2006] and often are *"lacking the capacity for innovation in the face of environmental change, which alters the initial conditions"* [CARNEY, 2012, p. 9.]

2. **Systems based on Psychosocial Aspects:** During the first management control revolution (1950-1980) the increased complexity of organisations meant that existing control techniques were not adequate enough and an extension towards to motivational factors became necessary for most of the new organisations. [CAPLAN, 1971] The complexity of new tasks required managers to persuade the members of the organisation to change historically passive behaviour and management not only to consider the formal aspects but also to include the behavioural aspects of the individuals influenced by the organisation and the organisational context in which they act. [CARNEY, 1997] Three key theories can be distinguished based on their approaches to human resources [AMAT, 1992]: the Human Relations School [ARGYRIS, 1990]; [HOFSTEDE, 1968]; the Trend based on human information processing [MACHINTOSH, 1981]; and the Open Systems Theory [LOWE AND TINKER, 1997]; [HOFSTEDE, 1968]. The common aspect of all of these theories is that they both influence and are influenced by people that form part of the organisation. Therefore, the control of individual behaviour may not only to be achieved by using quantitative measures but also by using psychosocial techniques. Furthermore, human beings are limitedly rational and tend to be satisfied with lower level gains rather than maximising them so the organisation's objectives can create conflicts with the individuals' objectives. [AMAT, 1992]

3. **Systems based on Cultural and Anthropological Aspects:** Since the second management control revolution (1980-1999) researchers found that [COOPER, 1983]; [DERMER AND LUCAS, 1986] control systems focusing only on the organisation's formal and psychosocial aspects are no longer able to explain the reality of the organisational structure and performance. With the rapid expansion of the large organisations, the anthropological aspects such as values, beliefs, stories or even celebrations through rituals [ALLAIRE AND FIRSIROTU, 1984] started to contribute to the achievement of the organisational objectives. Furthermore, as the economic environment has become more multinational and multicultural through the globalisation of business the culture as a key factor in control system has also been identified as a new control variable [AMAT, 1992]. Although such systems are considered advanced, the incorrect use can cause dysfunctions [CARNEY, 2012]. Some managers for example have mechanist perception on culture, which can lead to a belief that the culture can be easily manipulated and used as a way of solving management problems [MORGAN, 2006] (which can be harmful).

It is important to highlight that all three control systems detailed above still exist at different organisations depending on the size, activity and the management of the company. Furthermore, the combination of the different systems can exist at larger organisation, for example as part of the difference in management of divisions in matrix organisations. The economic environment and the economic challenges of different times have indirect impact, whilst legislations and the actual economic policies have a direct impact on the focus and the tasks of the control system. Before the recent (2007-2011) financial crisis the focus of control systems was on the information distribution [KÖRMENDI AND TÓTH, 2006]; [MANN AND MAYER, 1993] and the strategic support functions [TÓTH AND ZÉMAN, 2003] If we think about the root causes of the crisis, such as excess leverage, liquidity mismatch, conflict of bankers' interests etc. as the inquiry report of the United States [2010] highlighted, it can be stated that even though these aspects are all in the centre of management control, the importance of controlling was slightly overlooked leading to inadequate governance of the above mentioned factors [LENTNER et al., 2010]. The crisis impacted every financial institution around the globe [LENTNER, 2009] and forced them to make several changes to their management control systems.

Different institutions, due to their diverse strategies, had to make different levels of changes to their operations during the crisis. For example, Figure 4 shows two main paths that Northern European countries have followed since the crisis based on a recent study [RIKHARDSSON et al., 2016].



Source: Figure 2 and Figure 3 from CIMA study [RIKHARDSSON et al., 2016, p. 6.]

Figure 4: Financial crisis and changes in management controls in banks

According to the same study several controls have been enhanced[18] and the role of management control professionals "*gained increased importance with new external compliance requirements and changing emphasis of performance measurement systems and new values".* [RIKHARDSSON et al., 2016, p. 8.] The role of these professionals has been extended from design, operation and maintenance of management controls to interpretation (of concerns for example) and execution.

Following the financial crisis, financial management control systems around the world have been influenced by the changes in the regulations through the stricter accounting and auditing practices[19] and the overall re-regulation of the liberal money and capital markets. [LENTNER, 2010]. It highlights the fact that not only the local regulatory and supervisory bodies are responsible for sound controlling and compliance systems but also the financial institutions have social responsibilities to incorporate such systems into their operational practice. The stability of the financial system is common interest locally and internationally therefore the significance of bank management control, internal and external audit as well as central supervision has increased in the last decade. Especially the importance of the information provided through the products of the management control and accounting systems, such as internal and external reports, has been amplified. [LENTNER, 2011]

[18] Administrative controls (formalisation of controls through written policies), Rewards and Compensation, Planning, Cultural Controls, etc.
[19] Through the changes to the International Auditing and Assurance Standards (IAAS)

2.2 THE MANAGEMENT CONTROL SYSTEM

The practical manifestation of the above detailed management control concept and definition is the management control system at the organisation. The successful application of management control has two main requirements: a management control system or at least a collection of relevant procedures and principles; and the existence of management control culture at the organisation. An organisation has a unified management control system if the planning, budget versus actuals analysis and the decision-supporting information distribution systems are closely linked and all three are operated centrally or split only functionally (marketing-controlling; production-controlling etc.). [KÖRMENDI AND TÓTH, 2006] Other studies [DU et al., 2008]; [VAN IWAARDEN et al., 2006]; [BODA AND SZLÁVIK, 2001] also suggest the extension of these activities to execution (or the 'Do' element of the PDCA cycle[20]) or coordination [GAÁL, 2007] leaving only the decision making to the senior management. The second success criteria is the existence of the management control culture in line with the controlling function. In general, management control culture exists if the local and overall management activities of all core and supportive functions can be characterized by the following five principles and philosophies: feed-forward thinking (strategy and future oriented thinking); goal-orientation; focus on bottlenecks (continuous improvement of process); cost-orientation; decision focus. [TÓTH AND ZÉMAN, 2006] The following chapter details the functions, objectives, the strategic and operative processes along with the tools of the management control system. Author is also discussing the structure of the system and examines the specifications of the controlling at financial organisations.

2.2.1 The functions and objectives of management control

Current dissertation recognises four main management control functions: planning, supervision (comparison of actuals versus budget), coordination and information services. Although companies may put different emphasis on these functions, in general all four tasks are covered by the management control function. The above main functions have different objectives depending on the level of management involved (for example strategic management versus operative management):

Planning: "*Planning is the intellectual anticipation of possible future situations and the selection of desirable […] objectives and the determination of relevant measures that need to be taken*." [HORVÁTH, 2004, p.185.] With other words controlling planning is the identification and definition of the tasks that are required to be completed in order to achieve the objectives of the organisation. Depending on the main question, the level of management and the time horizon, management control planning can be divided into the following four levels

1) Corporate Policy: Getting the corporate policy right is vital for every organisation as this determines where the management want the company to be at the end of the long-term planning time-horizon. This corporate concept is recorded in the vision and mission statement.
2) Strategic Planning: The time horizon of strategic planning is 3-5 years depending on the industry. The strategic planning process starts with analysis of the environment external to the organisation (political, economy, competitors, customers and markets, suppliers etc.) and internal to the

[20] PDCA (Plan, Do, Check, Act) cycle, also referred to as Deming Cycle after Dr. W. Edward Deming, is an iterative management method primarily used to control and continual improvement of processes and products. [ROTHER, 2009]

organisation (resources, organisation structure, skills, comparative advantages etc.). This then used in conjunction with the corporate vision and mission to establish the system of strategic objectives which forms the basis of operative planning. [TÓTH AND ZÉMAN, 2006]

3) Operative Planning: Operative planning covers either mid-term (1-3 years) or annual planning. Operative plans arise from the strategic plan, which is transformed into objectives and budgeted values. Operative plans should be consistent with processes that influence the future of the organisation as these plans form a link between the strategic directions and the operative tasks.

4) Disposition: On this level of planning the organisation is aiming to prepare for unforeseeable events by constantly assessing the disruptive factors, their impact on the existing objectives and the probability of occurrence. Table 4 details the main questions, the required decisions and the plan elements of the different levels of planning.

Table 4: The levels of management control planning

Planning level	Main Question	Decision content	Plan elements
Corporate Policy	Who do we want to be?	Setup main objectives and constraints	Vision, mission statement, corporate concepts
Strategic Planning	Where do we want to go?	Find and select opportunities	Strategic Plans
Operative Planning	How do we reach our objectives?	Exploit existing potentials for success or build up new ones	Mid-term planning, Annual Planning
Disposition	How do we react to turbulences?	Take corrective actions to stay on course	Forecasting

Source: HORVÁTH, 2004, p.186.

Supervision (comparison of actual versus budget): The second main function of controlling is the supervision, with other words, the comparison of the budgeted metrics against the actual values. The key objectives of this function are: to check if (and how) the strategic and operative objectives are progressing; to provide decision makers with recommendations on intervention in case of a deviation from the corporate targets; and to improve the next planning cycle based in light of the new factors that caused the deviation in the first place.[21]

Coordination: The coordination function is sometimes merged into the information services or in some cases kept as a managerial responsibility. In more recently [GAÁL, 2007]; [HORVÁTH, 2004] coordination is considered as a separate management control function. The objectives of the coordination are: to link the sub-systems and sub-plans; to ensure there is co-ordination between the various areas and managers in their planning-, accounting-, reporting systems and even in their meeting (communicating the same vision, strategy and facts or with other words the "one version of truth"); and to support and promote comprehensive projects that come from directives, personal and departmental initiatives or through the regular planning process.

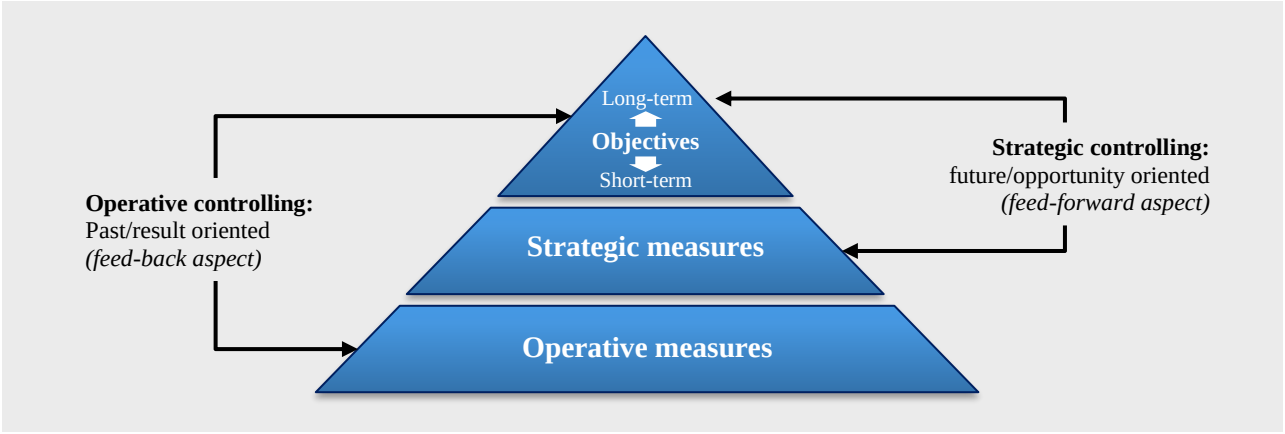
Information services: The fourth main function is the information services. Although, all of the functions detailed so far are closely related there are objectives and tasks that are unique to the respective function. Information services on the other hand is firmly integrated into all three activities, meaning that the same objectives are adapted to the relevant function. The three key objectives of information services are: to collect information for planning, supervision, coordination; to process

[21] A detailed review of Planning and Supervision can be found in Chapter 2.4

the information into output that supports the function it was processed for (e.g., market analysis for strategic planning or variance analysis for supervision); and to distribute the information in order to provide the adequate amount and quality of information with the optimal frequency to the respective decision makers.

2.2.2 The processes of management control: strategic and operative controlling

Due to the rapid changes in technology and economy the management of modern organisations are required to constantly monitor and re-evaluate the strategy for a long-term profitability. A need for a future looking, strategy supporting control system based on robust methodologies, like the ones applied in operational control, has naturally emerged amongst senior managers. As a result of this need for development, modern management control systems are composed by two fully integrated, methodologically standardised, hierarchically ordered, specialised sub-systems: the strategic management control and the operative management control sub systems. The above detailed functions and objectives for the two sub-systems are the same however both the input (type of information for example) and the output (type, quality and regularity of output for example) sides are specialised to the time-horizon and the management focus. Figure 5 demonstrates the hierarchy and the interrelations of the two sub-systems.



Source: Compiled by author using Figure 3 from [KALMÁR AND ILDZSA, 2012, p. 6.] based on Becker
Figure 5: The sub-systems of management control

Strategic management control sub-system: The strategic management sub-system is actively supporting the complex work of the strategic management. For example: the co-ordination function not only covers the objectives described in the previous chapter but it is also extended to the selection of external consultancy or the establishment of strategic and project teams. The following main areas are supported by the strategic management control [KÖRMENDI AND TÓTH, 2006]:

- The establishment of the organisation's vision (5-10 years development plan etc.)
- The definition of the organisation's mission (main objectives, operating policies etc.)
- The creation of the organisational culture model (key behaviours, organisational qualities etc.)
- The condition assessment of the organisation's strengths, weaknesses and opportunities
- The compilation of the system of strategic objectives (marketing, R&D strategies etc.)
- Supervision of the strategy implementation (feed-back aspect)

Since strategic management has been gaining popularity as an organisation control function [CHIKÁN, 1997] and new shareholder structures have been forming the need for information and a

new comprehensive performance indicator system became important. Strategic management control applies three key methodologies: financial measures of organisational value; management accounting performance indicator; and Balance Scorecard.

Operative management control sub-system: The operative management control is responsible for the annual (or everyday) controlling activity at the organisation by breaking the strategic plans down to operative actions. The focus is on the effectiveness of the processes, tactical plans, short term profitability/rentability and costs. This subsystem is also constantly monitoring the progression towards the organisational objectives. Any deviation from the plan (driven by internal or external changes) are reported and fed-back to senior management and will also form part of the next operative and strategic planning cycle. The following main areas are supported by operative management control:

- Setting the mid-term/annual objectives of the organisation
- Identification of production or process bottlenecks
- Establishment of the cost-structure (cost-, profit- and service centres)
- Performances supervision
- Comparison of budget and actuals.

2.2.3 The tools of management control

In order to operate the above detailed management control system effectively, controlling functions require a set of robust tools, methods and processes. In the current chapter author introduces the most common management control tools based on existing literature [KÖRMENDI AND TÓTH, 2006]; [BODA AND SZLÁVIK, 2001]; [ZÉMAN AND BÉHM, 2016]; [LENTNER, 2013]; [FRANCSOVICS, 2005]; [SZÓKA, 2007]; [HORVÁTH, 2004]; [OTLEY, 1994]; [CHENHALL, 2003]; [MERCHANT AND VAN DER STEDE, 2011] keeping in mind that variations, combinations and other methods may exist in the industry practice.

Management Accounting: Management accounting is probably one of the most integrated tools of management control. In Europe, for example in the United Kingdom management control and management accounting are often used as synonymous terms unlike in the German or Hungarian practice where Management Accounting is an inevitably important, supporting element of management control. As we have seen on Figure 3 Management accounting was developed in 1950s (and after 1991 in Hungary with the introduction of IAS[22]) when the need emerged from the managers for a robust but flexible accounting information system, which was not only able to provide the formal, regulated external accounting information through the profit and loss accounts and tax returns (provided by financial accounting) but also supported planning, analysis, supervision and internal information management at the organisation. This new accounting system has become a primary tool for management control, especially for the strategic control sub-system. The latter sub-system is supported by a range of strategic management accounting indicators:

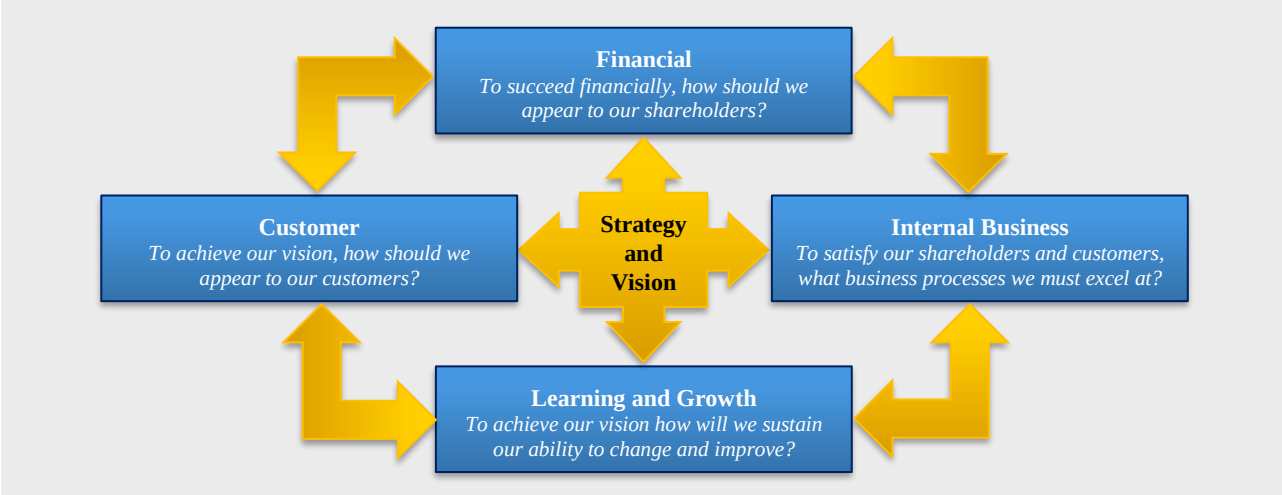
- Value based financial indicators (such as Shareholder Value or Discounted Cash Flow)
- General organisational indicators (the various elements of the Balance Scorecard)

[22] IAS: International Accounting Standards, for more information please visit: https://www.iasplus.com/en/standards/ias

Management Accounting is also providing insight to operative controlling through detailed cost-plan calculations and cost-coverage-income ratios like static and flexible cost calculation; standard costs; activity based cost and process cost calculation; direct costing, etc.

Organisational tools and methods: Management control is an adaptive system which means it is constantly evolving as new challenges are rising. It is flexibly adopting the latest results and tools of other interlinked organisational and management sciences in order to support the management with fast reaction. The followings are only a few advanced tools applied by management control:

- Benchmarking is an advanced, strategy supporting method, which helps the management to identify internally (within the organisation) or externally (at similar company from the industry for example) the best practice, the base point which the organisation is comparing the performance, processes or structure against, with the ultimate aim of adopting the "best in class" procedures and re-engineering existing processes or the organisation itself.
- Scenario building technique enables management control professionals to parameterise internal and external economic factors in order to build forward looking scenarios. The scenarios are built with a cause and effect approach, which helps the strategic management to visualise the path to the predicted vision.
- SWOT-analysis is primarily used for strategic planning. It helps management to understand the interrelations between the internal and external environment of the organisation by analysing the Strengths and Weaknesses of the organisation with the external Opportunities and Threats. It helps strategic management to identify potential risks and the areas where the company has the highest possibility to succeed.
- Balanced Scorecard is a harmonic key performance indicator and objective planning system balancing the focus between financial and non-financial indicators. [KAPLAN AND NORTON, 1992] As Figure 6 demonstrates balanced scorecards have four main aspects: Financial (How do we appear to shareholders?); Internal Business (At what processes should we excel?); Customer (How do our customers perceive us?); and Learning & Growth (What should we learn to grow and prosper?).



Source: Compiled by author based on [CSATH, 2004, p. 271.]

Figure 6: The different aspects of Balance Scorecard

Within each aspect, strategic management sets up a range of objectives, key performance indicators, target values and detailed actions that move the organisation towards its vision.

- Business Process Reengineering (BPR) is primarily used by the operative control sub-system. The method targets inefficient processes and potential bottlenecks with the aim of optimisation and re-organisation.
- Outsourcing is a drastic way of modernisation at the organisation. It mainly targets non-core activities that are costly and inefficient and could potentially be completed by another organisation external to the company, cheaper and more efficiently without damaging the business integrity.

Information Technology: Probably it is not emphasised enough in the literature how much an organisation is reliant on information technology in modern days. In the world of the 'internet of things', companies have access to a wealth of financial and non-financial information almost real time from internal and external sources. With the latest development of API[23] technology it became easier for organisations to acquire market intelligence data Furthermore, with the cheap and efficient processing the commercialisation of such data is available for a wide range of companies.

Most of the small and almost all medium to large organisations should capture, store, process, analyse and commercialise data on a daily basis. Depending on the size of the company, the information requirements and sophistication of the system, companies use one or the combination of the following solutions: a local controlling software (in some cases as part of an accountancy software package); an integrated management solution; or a cloud based solution. These modern software packages are deployed at the final step of the implementation process as an effective control module is linked to all functional (marketing, logistic, production) modules. The new cloud based solutions are generally flexible and scalable therefore new modules and data sources can be added efficiently. Such systems at modern organisations are used at all management control functions from planning to information distribution. For example, these solutions can automate the monitoring and most other elements of the supervision. Current research was triggered by the rapid development of information technology and data science, which now offers a wide range of statistical methods efficiently accessible by traditional management control areas. A detailed review of information technology and data science as key enablers of management control method improvement can be found in Chapter 2.5.

2.2.4 The management control at the organisation

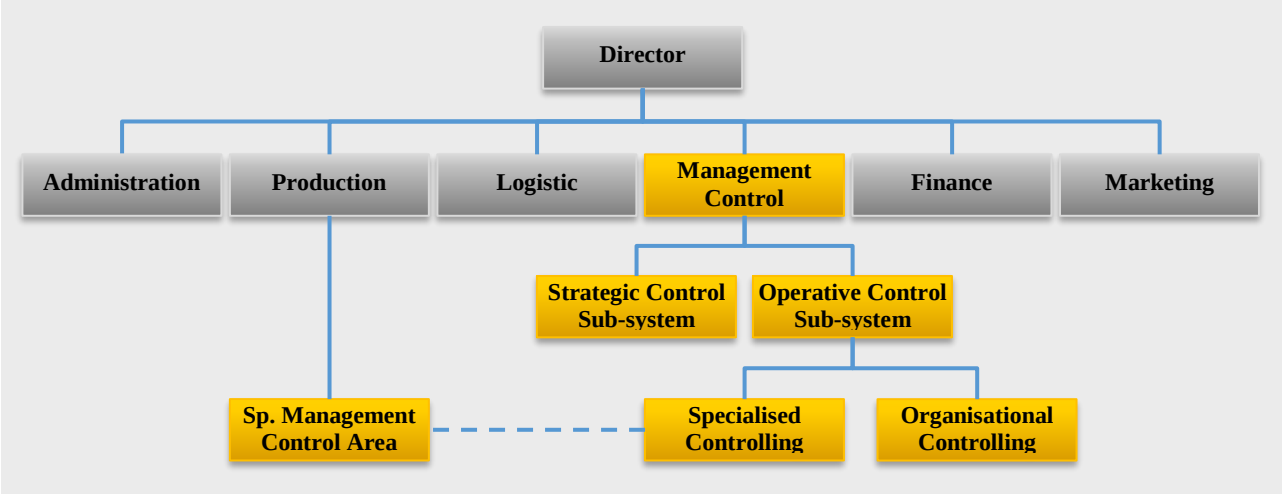
While the scientific community of management control is aiming to generalise the concept and standardise the different methods, the practical management control within the different organisations is organically heading towards to the specialisation and differentiation of the control system. Management control can be adapted at any organisation where value adding processes are in place. Therefore, the classical management control concept was adapted at manufacturing, trading, service and banking organisation at first. However, in the last three decades researchers have registered an increased adoption rate by the governmental bodies, state institutions, civil organisations, national healthcare and even by the Church. [KÖRMENDI AND TÓTH, 2006] The adoption of the concept also means the adaptation of the system, functions and methods to the specific industry. However, the differences of the industry, or external environment is not the only driver of specialisation. The increased complexity of the organisational structure and activities naturally created a need for the senior management to have increased focus, visibility and control over key areas and processes.

[23] Application programming interface

These, newly emerged, integrated sub management control system areas are called specialised management control areas. Depending on the organisation and the specific processes targeted by the specialised controlling, the organisation can have various specific functions. The following list contains the most common areas:

- Project management control area (overall project supervision from initiation to closure);
- Eco management control area (supporting environment focused management functions);
- Quality management control area (total quality management support functions);
- Logistic management control area (the control of logistic management performance);
- Marketing management control area (marketing and product management control);
- Human Resource management control area (human resource management and performance);
- Financial management control area (classical backward looking accountancy based area).

Regardless of the number of specialised controlling areas, probably the most important element of the controlling system at the organisation is the management control administration (team, group or department) and its members, the management control professionals. In general, small companies do not require separate control functions. As long as the management control philosophy and concepts are applied in the day to day management activity, most of the management control tasks can be completed by the company accountant. At the same time, medium and large companies will certainly benefit from a dedicated management control function at the organisation. Figure 7 demonstrates how the central and specific management control areas are integrated into a functional organisation structure. It is important to highlight that although a specialised management control area can exist under an organisational function, it should always be joined up with the central management control function in order to keep consistency. [ZÉMAN, 1998]



Source: Compiled by author

Figure 7: The management control function at the organisation

It is also advised to keep the central management control function independent from all other departments. Although, it is closest to the Finance and Accountancy function, the objectives and the information system of management control are separate from this function. Author recognises that different organisational structures require different management control structures, for example controlling of divisional organisations, company groups [TÓTH AND ZÉMAN, 2003] or even hospitals [TÓTH, 1998]. It is important to establish the roles, responsibilities and the authorisation of the central management control function and its controllers so they can effectively can do their job.

The main role of the management control professional is to provide transparency on business performance and costs to all levels of the management from strategic planning to the operative realisation. It is also the responsibility of the controller to establish and maintain a management control system that satisfies the needs of the management and the requirements of a control system design. With other words, the system should be in line with the objectives and strategy of the organisation and with the responsibilities within the structure. In addition to the above detailed, the following list shows the key tasks of the management control professionals [HORVÁTH, 2004]:

- Co-ordinate sub-targets and sub-plans in a holistic way;
- Organise a system of reporting, which is both forward looking and covers the organisation;
- Moderate the management control process so that every decision maker can act in accordance with agreed objectives;
- Ensure that managers receive all necessary information with a timely manner
- Act as internal consultants to all decision-makers for a wide range of questions (in planning, performance, accounting, finances etc.)

Management control professionals, in order to effectively perform these tasks are required to be educated in economics, organisational studies, finance and accountancy, ideally to be experienced in mathematical-statistical methods and to be proficient computer users. In addition to the professional skills controllers should also have strong analytical approach and creativity. They should be able to work in a team, open for new knowledge, should be a strong influencers, co-ordinators and motivators with a good sense for diplomacy and strategic thinking. [KÖRMENDI AND TÓTH, 2006] Based on the mix of these qualities and their primari role within the management control system, controlling professionals can be clustered into three groups: registrar (classical accountancy-oriented); navigator (future and event oriented); innovator (management and system oriented). [HENZLER, 1974]

2.2.5 The specifications of Bank management control

The management control of financial institutions[24] applies the same forward looking and goal oriented philosophy, principles, basic function (planning, supervision, coordination, information services) or even the basic structure (strategic and operative control) as the management control of non-financial organisations. The key differences come from the different nature of activities a bank is doing compared to other organisations. This means that tasks and objectives of the functions and activities covered by the sub-systems also include bank specific elements. The main tasks of bank management control are the following [ZÉMAN et al., 2013]:

- Improving accounting systems towards management accounting;
- Strategic and operative planning;
- Supervision (cost, income, risk analysis, comparison of actuals and budgets, trend analysis);
- Information services: development of a comprehensive managerial information system, reporting system with an appropriate IT support.

One of the main specifications of the management control function at financial organisation is the distinction between the management control tasks of central functions and the tasks of distribution network (decentralised functions). Table 5 highlights the centralised and decentralised tasks. It can

[24] such as banks, investment and insurance providers

be stated that both sets of activities should be synchronised and delegated to an optimum level where the central management functions still have sight of all activities but the principles of management control are embedded to the decentralised management functions.

Table 5: Centralised and decentralised bank management control tasks

Centralised Tasks	Decentralised Tasks
Development of a decision oriented evaluation system	Development of decision focused culture
Centralised cost and expenditure planning	Decentralised cost planning
Development of business policies	Decentralised (delegated) planning tasks
Evaluation of target and actuals	Decentralised (delegated) control tasks
Development and maintenance of the management information system	Regulation of decentralised information processing
Establishment of an integrated target evaluation system	
Profit requirement analysis	
Profitability and risk analysis	

Source: Compiled by author based on [BAGINÉ et al., 1999]

The Strategic bank management control sub-system: in addition to the original strategic management control responsibilities, bank management control is also focusing on the development (e.g., growth potentials), structural (e.g., structural risks and yields) and security (e.g., market risks) questions at the organisation. It covers the following additional fields:

- Portfolio management: the right financial product mix enables financial institutions to achieve their strategic and profit objectives by reinforcing their market position. Portfolio management is the bank specific strategic management control function that is responsible for the establishment and maintenance of the product strategy. It covers three dimensions: 1) Offered product and service range (current account, loans, savings, investment etc.); 2) The product and service breakdown with the specifications; 3) Service and product quality (security, discretion, speed, etc.).
- Structural balance sheet management: Structural balance sheet management is another management control area unique to financial institutions. The key objective of this function is to establish a long-term balance between the risks and profitability by managing the structure of the balance sheet. It is achieved by the constant review and evaluation of risky items (size, probability of default, time horizon etc.) compared to the interest and other income sources.

The Operative bank management control sub-system: Similarly to corporate controlling, operative bank management control is responsible for the breakdown of the overall strategic goals. The key tool of bank management control is the, so called, budget management which disintegrates and distributes the strategic plan to mid-term and annual plans by the specific organisational areas and sub-functions. Budget management also involves supervision (comparison, analysis and evaluation of budged against actuals) and the preparation of management decisions. Operative bank management control also covers: operative planning; liability and asset management; industry analysis and insight; risk analysis; investment management; and distribution network management. It is worth to mention that in the banking industry the activities controlled by the operative management control are not always arranged vertically (by bank operations and areas) but also horizontally (profit, cost, service centres). [ILDZSA, 2009] This is particularly important when a bank is using Activity Based Costing method. [MABBERLEY, 1999]

Customer level bank management control: as part of the operative controlling, the banking industry specific customer level management control activity plays a vital role in the prudent operation. This activity is integrated into the strategy and regulated by the governmental and other official regulator bodies (such as Central Bank, Basel Accord etc.). Based on the strategic direction, it is the task of operative controlling to regulate, supervise and coordinate information between the relevant departments (e.g., risk, finance, etc.). It covers the following two main tasks [SÁGI, 2007]:

- Credit Rating: The main aim of credit rating is to identify potential risks or probability of customers cannot repay the loan (defaulting). The credit rating procedure and the new customer portfolio are constantly monitored and adjusted based on the changes of the economic environment. Credit cut-off scores are periodically evaluated. The credit rating regulations are derived from the financial institution's market strategy, with other words from the 'risk appetite'. [KALMÁR, 2010] The main aim is to find the balance between risk, income and the strategy. Credit ratings are used at planning and budgeting as well as at the supervision and information management functions.
- Monitoring: The monitoring starts with loan approval and ends when the last repayment amount is credited or written off. Monitoring helps operative management control professionals to identify any potential issues the with the loan portfolio (e.g., by monitoring expected probability of default against actual loan defaults) which would potentially require the adjustment of credit acceptance scores or any other management decision.

As we have seen in this chapter management control has a vital role at the organisations. The forward looking, decision oriented management control on both, strategic and operative levels support banks and non-financial organisation to improve profitability and reduce bottlenecks and costs by helping managers to make coordinated and informed decisions. Moreover, it also supports the overall organisation to adapt to the everchanging market environment. In the next chapter author is going to describe how the market and customer oriented marketing management and the future and goal oriented controlling are integrated into the marketing control specialist management control area and how organisation utilise and benefit from this function.

2.3 MARKETING CONTROLLING

In the second half of the 20th century the increased competition on many markets meant more options to choose from for the customers. These customers required information to decide what products or services to purchase. Companies started to realise that providing information in the form of advertisements, packaging, promotions and other customer communications helped to increase their profitability. Organisations with the aim of staying in (or ahead of) competitions started to adopt the marketing concept[25] and the marketing management practice[26]. Since the late 1970's it has become evident that marketing is playing a vital role in the success of a company. The objectives of marketing management have increased and started to cover activities ranging from customer and market analysis through the selection of the right product and promotion mix until the establishment of the distribution strategy (even including the selection of distribution channels). [KANDIKÓ AND LEHEL, 2012]

[25] Marketing is about identifying and meeting human and social needs. One of the probably shortest, yet complete definitions is saying that marketing is "meeting needs profitably." [KOTLER-KELLER 2012, p.5]
[26] customer needs should be satisfied in a better and more attractive way

Since marketing management, *"the art and science of choosing target markets and getting, keeping, and growing customers through creating, delivering, and communicating superior customer value"* [KOTLER-KELLER 2012, p.5] became a standalone scientific area, the number of researchers and marketing professionals has increased. The increased talent pool and the fast spread of consumer technology (TV, mobile, smart technology, etc.) created multiple opportunities and solutions to efficiently meet customer needs profitably. However, "*this orientation towards marketing has important strategic and managerial consequences for banks"* [CLARKE et al., 1988] and for non-financial organisations. In the last two decades with the increased marketing costs and the difficulties around the quantification of marketing-cost effectiveness, the need for an independent planning and cost controlling function focusing on efficiency, management decisions and future, has emerged amongst senior managers. [FARRIS, et al., 2008] Although the 'marketing audit' and 'marketing cost control' concepts appeared in in the 1980s, the scientific literature of the new phenomena, called Marketing Management Control (or marketing controlling) started to be researched and well documented only in the last two decades.[27] Marketing Management Control is amalgamating the customer and market focused Marketing Management with the goal and future oriented Management control in support of the organisational strategy and the decision makers of the organisation.

2.3.1 The Definition of Marketing Management Control

Similarly to the definition of the management control, the definition of marketing management control is also inconsistent in the literature. Different authors emphasise different elements of the marketing control in their definition. Definitions are clustered around the following three key approaches [HAJDÚ, 2016]:

- **The classic marketing management control approach**: Authors representing this approach [KEINER, 1980]; [HORVÁTH, 1986]; [FOSTER AND GUPTA, 1994]; [BÖCSKEI, 2006]; [KANDIKÓ AND LEHEL, 2012]; [SZŰCS, 2014] see marketing controlling as an integrated element of the management control system, responsible for the strategic and operative planning and supervision of the organisational marketing.
- **The marketing coordination and supporting approach:** A group of researchers [JÓZSA, 2006]; [BRÁVÁCZ, 2008] uses marketing as the starting point and identifies marketing controlling as a marketing function. They see the key coordination between the financial and marketing plans, cost plans, actuals and budget as the primary function supporting the marketing activity at the organisation.
- **Marketing information services approach:** More recently professionals [REICHMANN, 2001]; [KÖHLER, 2006]; [HAJDÚ, 2016] have highlighted the information services function of marketing control, identifying the synchronisation of the marketing information system with marketing planning and supervision as the main objective.

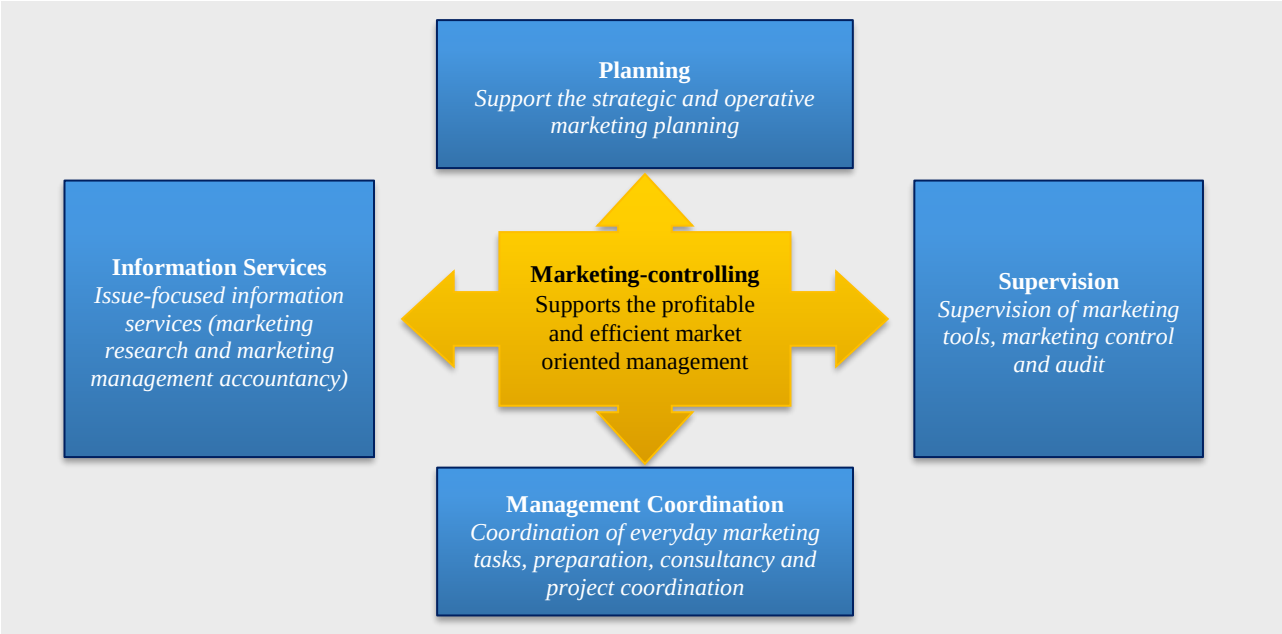
Current dissertation recognises the following definition of marketing management control: Marketing Management Control is a specialised area of the Management Control System. It covers a collection of marketing system specific strategic and operative processes, including: marketing planning (targets, costs, etc.); supervision (business analysis and management reporting, actual and budged

[27] Author has conducted a hypothesis testing, as part of current dissertation in Chapter 4.1 to determine whether Marketing Management Control is a new phenomenon, emerging from the literature in the last two decades or it has always been a specialist area of management control.

comparison, etc.); coordination within marketing management and between the central management control function; and information services (synchronisation of the marketing information system with the marketing planning, supervision and the central management control information system). The aim of Marketing Management Control is to help the organisation to achieve its strategic objectives by managing and providing the adequate level and quality of required management information on the marketing activity; creating transparency for all stakeholders; and supporting the decision-making process of the management.

2.3.2 The functions and objectives of Marketing Management Control

The marketing management control has the same four main functions adapted to the marketing environment, as the central management control system. Figure 8 shows the system of functions and objectives assessed in the current dissertation.



Source: Compiled by author using Figure 3 from [HAJDÚ, 2016] based on [KÖHLER, 2006]

Figure 8: The functions and objectives of marketing controlling

Marketing Planning: As part of the marketing planning function, marketing controlling is responsible for establishing the performance standards (objectives, norms, indicators) based on the agreed marketing strategy. The objective of this function is to set well-defined standards and to be the link between expectations and performance as well as to ensure the indicator system is recognised by the relevant stakeholders for effective performance management. [KANDIKÓ AND LEHEL, 2012]

Marketing Information Services: The link between marketing planning, supervision, management coordination and in fact the central management control system is the information services function, which records all required information, serving the other functions and the management directly. The objective is to establish and maintain a decision and issue oriented, integrated to the central management control system or information management service to facilitate solutions and fast management reactions. [KÖHLER, 2006, p.43.]

Supervision: Similarly to the mainstream management control, the supervision function covers the comparison of actual performance against plans (or targets). More specifically, marketing supervision is composed by four key audit areas: process audit; strategic audit; marketing-mix; and marketing organisation audit. This information is used for feed-back and for future actions (feed-forward information for strategy change or next planning cycle). The objective is to understand the deviations from the plan, to uncover the root-cause and advise the management on possible corrective actions in order to reach the strategic aims.

Management Coordination: The management coordination function is strongly related to the supervision and information services, as the outcome of these functions requires the coordination of separate management functions. [WEBER AND SHAFFER, 2006] Furthermore, marketing management control should ensure that decisions are in place and coordinates the execution in collaboration with decision makers. This function also covers consultancy support to projects and the preparation of other strategic activities.

2.3.3 Strategic and operative marketing management control

Following the logic of Chapter 2.2.2 we can see that marketing controlling can be arranged into two sub-systems based on the time horizon such sub-systems support. Strategic marketing control supports the long-term marketing management activities, it also enables organisation to adapt to the changes in the economic environment and to identify the long-term success potentials. Operative marketing management control operates on mid- and short-terms and supports operative planning and analysis with the focus on short-term success factors. Both sub-systems have marketing management area specific features compared to the central management control sub-systems:

Strategic marketing management control: Strategic marketing control is organisation-oriented with a simultaneous focus on the operating environment. The level of planning is strategic with the management objectives of growth, sustainability, comparative advantages and the identification of success factors. Strategic marketing management professionals often use comprehensive internal and external analyses to identify the opportunities, threats, strengths and weaknesses of the company. [HORVÁTH, 2003]

- Strategic marketing planning is focusing on the long-term existence of the organisation, it operationalises the objectives and assesses the feasibility.
- The strategic marketing information services function is closely linked to the supervision function in order to identify deviations from the original marketing strategy and to provide alternative plans containing corrective actions in response to internal and external changes.

Operative marketing management control: Operative marketing control focuses on the efficiency and profitability of marketing plans. The level of planning is at operative and tactical levels with the management objectives of corporate and marketing profitability, efficiency and solvency. The primary source of management information is the complex system of indicators based data extracted from the accounting and management information systems. Operative marketing control is the subordinate system of the strategic marketing control applying all control functions:

- Operative planning derives annual plans from the strategic sub-system, creates the annual marketing cost plans based on the organisational structure and marketing activity plans.

- Supervision has an important role in operative marketing management control by constantly analysing differences in actuals versus marketing targets and acting as an early warning system for managers.
- As part of the coordination function the sub-system is mainly responsible for the synchronisation of the marketing tools and mix. [HAJDÚ, 2016]

2.3.4 The tools and metrics applied by Marketing Management Control

The tools used by the marketing management control system can be categorised in two different ways: 1) clustering tools by two dimensions: by function (marketing planning, coordination, supervision, marketing organisation) where the tools are applied; and by the subject of analysis (product, product-mix, markets, customers, competition) [KÖHLER, 2006]; 2) grouping based on the application system (strategic marketing control tools, operative and general marketing control tools) [REINECKE, 2006]. Table 6 shows the management control toolkit based on the first option.

Table 6: The system of marketing management control tool

	Marketing Planning	Marketing Organisation	Coordination	Marketing Audit and Supervision
Products	Industry portfolios Positioning studies Product lifecycle planning Investment calculations	Gross margin calculations Production cost planning Brand value estimation	Establishment of the compensation system of distributors	Product level transfer price Product level process cost
Product-mix	Response prediction Activity based costing Decision calculation	Functional cost planning	Detailed preparation for contract negotiations	Media campaign evaluation Logistic system control Marketing-mix audit
Markets	Sales volume potential forecasting models Distribution network optimisation models	Gross margin calculations by market type Regional cost planning	Establishment and distribution of regional marketing objectives	Intermediary gross margin calculations Strategic audit of markets
Customers	Future client value estimate Customer portfolios Customer segmentation	Customer relationship management cost planning	Evaluation of customer satisfaction	Customer gross margin and cost calculations
Competitors	Early indication of strength and weaknesses with SWOT analysis	Competitor profit and loss analysis	Benchmarking for motivation	Benchmark of marketing audit indicators

Source: Compiled by author using Figure 6 from [HAJDÚ, 2016] based on [KÖHLER, 2006]

As mentioned above the tools and measures can also be categorised by the applying sub-system:

- **Strategic marketing management control supporting tools:** Strategic marketing control tools tend to be qualitative measures, like scenario analysis or benchmarking and even 'hard measures' like ratios and indicators are converted to softer, balanced scorecard type metrics. These are the most common tools used by strategic management control: early warning system; industry structure analysis; strategy benchmark; benchmark of strengths and weaknesses; segmentation (customer, market, competitor); investment calculation; brand and customer value calculations; long-term cost planning; control and supervision of basic marketing activities, etc.
- **Operative marketing management control supporting tools:** Operative marketing control has a more metricised approach as the key information source is the financial accounting and marketing information system. It is using a system of indicators, gap analysis, information management services to marketing functions based on market insight, sales and finance data; information for the marketing-mix composition; short-term cost planning and audit; customer satisfaction and complaints analysis; pricing; distribution and service audit, etc.

- **General marketing management control supporting tools:** Finally, there are other tools mainly related to the coordination function that are integrated into both subsystems. The range of tools and processes includes the establishment of the compensation and motivational systems; target cost calculations; marketing and sales project analysis, planning and supervision; evaluation of marketing contracts; knowledge management of marketing and sales management areas.

As we have seen in Chapter 2.3 marketing management control is a specific management control area with similar, but marketing specific responsibilities, tools and objectives compared to the central management control system[28]. Author in the next chapter is focusing on the planning and supervision functions especially the existing planning and evaluation methods of the central, marketing specific and bank specific management control systems to see if there is an opportunity for improvement.

2.4 PLANNING AND SUPERVISION IN CONTROLLING

In the current chapter, building on the details introduced in Chapter 2.2.1, author is reviewing existing planning and supervision practices based on the literature. Furthermore, author is drilling down a level to cost specific planning and evaluation tools and calculations as well as the most commonly used marketing cost related measures and techniques.

2.4.1 Management control planning in practice

The basis of effective planning is the system of goals and objectives defined by the management based on the external environment and the internal resources of the organisation. Goals can be ambitious however they should be realistic and achievable at the selected timeframe. Planning is responsible for defining tasks and ensuring that conditions are adequate to achieve the goals. Management control planning has three dimensions:

1. **Workflow type dimension**: there are three widely used methods in this dimension: bottom-up planning method (when the final plan is the aggregate of the unit level plans); top-down planning (when the master plan is broken down to unit level plans); and mixed approach.
2. **Time horizon dimension:** based on the time horizon of the plan we differentiate the two ends of the dimension: strategic planning (3-5 years); and operative planning (0-3 years).
3. **Subject area dimension:** this dimension is split by the specific economic and management area that the actual planning phase is targeting: investment planning; business planning; strategy, sub-strategy and tactical planning This dimension is strongly related to the time horizon[29]; performance planning; cost and income planning; cash flow planning etc.

The following list contains the main management control planning activities and their details based on the above described three dimensions [KÖRMENDI AND TÓTH, 2006]:

Investment planning: It covers the planning of a wide range of strategic activities such as acquisitions, investments, the change in company value, the change in the return on equity and shares etc. It relates to long term decisions and most often used with a “top to bottom” workflow. The most

[28] Introduced in Chapter 2.2
[29] For example: investment planning is typically a strategic time horizon activity unlike cost planning which is more likely relate to short term planning, etc.

common tools and methods applied during investment planning are the Balance Scorecard; Shareholder Value and Value Based methods.

Strategy and sub-strategy planning: The strategy planning from the management control perspective generally means the establishment of key performance indicators for the goals that the organisation is aiming to achieve in strategic term (for example: sales increased by 20% at end of the third calendar year, etc.). The time horizon is typically strategic and generally "top to bottom" method is applied. The sub-strategies (or the action plans) are still covering strategic time frames however are also containing actions and projects for the specific area (R&D, developments, investments).

Tactical planning: The tactical plan breaks down the investment, strategy and sub-strategy plans by year for the total length of the plans. It defines the key figures (budgets, targets, overall cost, etc.) per annum, which forms the basis (or input) of the operative planning.

Cost planning is one of the key activities of management control planning. However, due to the detailed review required for the current dissertation and to avoid duplication, it is discussed in details in Chapter 2.4.2.

Performance planning: The performance planning is typically an operative planning procedure mainly applied to establish the annual performance and turnover plans. The sales volume of products and services calculated based on capacity of the value adding processes and the market demands, forms the basis of the performance calculation. The annual sales volume (or performance) is weighted with the unit market price to calculate annual turnover. Depending on the ability of the management control system to breakdown costs to lower level cost centres, performance planning can be used to identify bottlenecks and process issues by calculating average lead time against available production capacity. Performance planning is often a 'bottom to up' workflow.

Result Planning: The final planning stage of operative planning at the profit-oriented organisations is the result (or break-even cost) planning. This is primarily a 'top-down' workflow using the unit cost (established during the cost planning activity) and the total planned turnover (calculated as part of the performance planning) to determine the operational profit. Depending on the level of planning, the management control planning can reveal the proportional profit contribution of organisational units or profit centres.

Financial and cash flow planning: Financial or liquidity planning is an additional planning element of cost and performance planning used for cash flow forecasting. The main objective of this planning function is to forecast income and spend flow throughout the calendar year to determine if (and when) external funds are required.

2.4.2 Cost planning methods

Cost planning is probably the most often used element of the operative planning procedure which is used to determine the total costs and inputs for a financial year. Author is reviewing the main cost planning procedures with the most common methods based on professional literature [KÖRMENDI AND TÓTH, 2006]; [BODA AND SZLÁVIK, 2001]; [HORVÁTH, 2004]. Körmendi and Tóth (2006) identified four main practical types of cost planning: Classical cost planning, process cost planning, activity based costing and target costing.

Classical cost planning: According to the classical cost planning method, cost should be planned based on where the cost is occurring (cost centre) and for what product or service (cost objects) they occur for. Therefore, in case of value adding processes the efficiency of such cost planning depends on the ability of the controller to break down and assign the costs to cost centre and cost objects. It is reasonably straightforward at manufacturing companies but institutions and service provider organisations are often forced to use other methods. Classical cost planning methods are typically bottom-up workflows, starting with the establishment of the planned costs for cost centres then building up the unit cost structure of cost objects. Based on the product structure and the planned production or sales volume the financial controller can calculate the total variable cost of the company. At the final step fix costs are added to calculate the total cost at the organisation. In order to calculate fix costs or even to determine specific cost lines within general or unit costs, classical cost planning applies the following methods [SZÓKA, 2007]:

- **Base budgeting:** This method is primarily relying on actuals from the previous planning cycle. The aim is to preserve the value of the assets on real terms. This method provides very limited flexibility on changes to the plan throughout the economic year.
- **Zero base budgeting:** This method does not take the actuals or the plan of the previous planning cycle into account. It starts planning from scratch (or zero). The management of the cost and profit centres are responsible for producing the respective cost plan, which is collected and rolled up to organisational level. This method is time consuming however excludes planning issues of the previous planning cycles.
- **Top-down budgeting:** In terms of cost planning, the top-down budgeting method is based on annual cost plans determined centrally and cascaded down to the different units. Due to the centralised nature of it, this method is inflexible, planning errors are generally solved with resource reallocation from one unit to another.
- **Bottom-up budgeting:** This method, similarly to zero base budgeting, requires local or unit managers to compile their annual cost plans based on their respectively chosen methods. The advantage is that units are closer to the market and therefore produce more accurate plans however units could have different interests and development plans which can lead to inconsistent strategy.
- **Mixed budgeting:** Depending on the organisational size, mixed budgeting is one of the most used classical methods. This is a combination of the top-down and bottom-up budgeting, which means headline cost numbers are defined centrally then cascaded down and challenged during the planning period. Coordination between central planning and the units is vital for an effective execution of this method.
- **Rolling forecast:** The latest classical method is the rolling forecast which requires monthly re-planning but offers the most flexibility and accuracy when compared against the rest of the methods. The method is based on a 12-month plan but instead of a one time annual planning period, there is a monthly evaluation process of the plan for the next 11 original, plus a 12^{th} (new) month. This method requires continuous evaluation of the strategy, market and capacities in order to update the different parts of the different financial plans (P&L plan, Cash-flow, Performance plan, etc.)

Process cost planning: Process cost planning is generally applied at organisations where the classical, cost centre and cost object oriented planning is unable to provide adequate management information for strategic or operative decisions on production due to the nature of the production

activity. For example, mono-product companies with closed production cycle often split the production to different phases, which makes cost assignment to different cost centres difficult. Typically, these large organisations, large agricultural companies and retailers apply a process oriented management structure. The cost planning should align to this. From a planning perspective, the accurate definition of the start-to-end, main and sub-processes (based on the different phases of the production) is vital. As part of the process definition the financial controller has to determine the right contribution (proportion) of the supporting functions to the main process. This schema is then applied for the planned production activity to determine annual costs.

Activity based costing: Activity based costing (or ABC) is a specialised version of process cost planning, applied mainly by large (primarily service provider) organisation, such as financial institutions. [MABBERLEY, 1999] The aim of the method is to rationalise resources used by the organisation to achieve its goals through information provided to decision makers about the resource and cost required for production, customers or even for certain transactions. The method can only be applied with a well-integrated information management system as good measurements and extensive data are vital. This method enables the management to identify bottlenecks, inefficiencies, non-profitable customers or markets. There are three practical ABC methods in the literature [BARRETT AND HAWTHORN, 2005]:

- **Time-splits ABC method:** This is the simplest of the ABC methods. Managers are required to record and split the time spent on different activities[30] by their resources. This information then forms the basis of cost allocation and assignment to resources and ultimately to activities. The benefit of the method is that it does not require additional information when combined with the existing accountancy data, therefore can be used for quick organisational modelling. However, the actual recording activity requires significant manual labour.
- **Time-capture ABC method:** This method is used to primarily determine how resources are split between projects and customers. The main application area of this is at development research of professional services organisations where the tasks are repetitive and difficult to categorise or phase them. The method is unable to highlight excess and unused capacity. Furthermore, if it is used for internal billing it can cause dissent among staff.
- **Time-driven ABC method:** *"Time-driven costing is where cost driver rates are based on the practical capacity of the resources supplied, measuring or estimating the amount of time taken to perform an activity."* [BARRETT AND HAWTHORN, 2005, p.8.] Although the method is addressing elements of the excess capacity question unless the data is readily available, robust and reliable, the time-driven ABC method can generate more problems than it solves. Due to the granularity required for the model, the method can generate large amount of data.

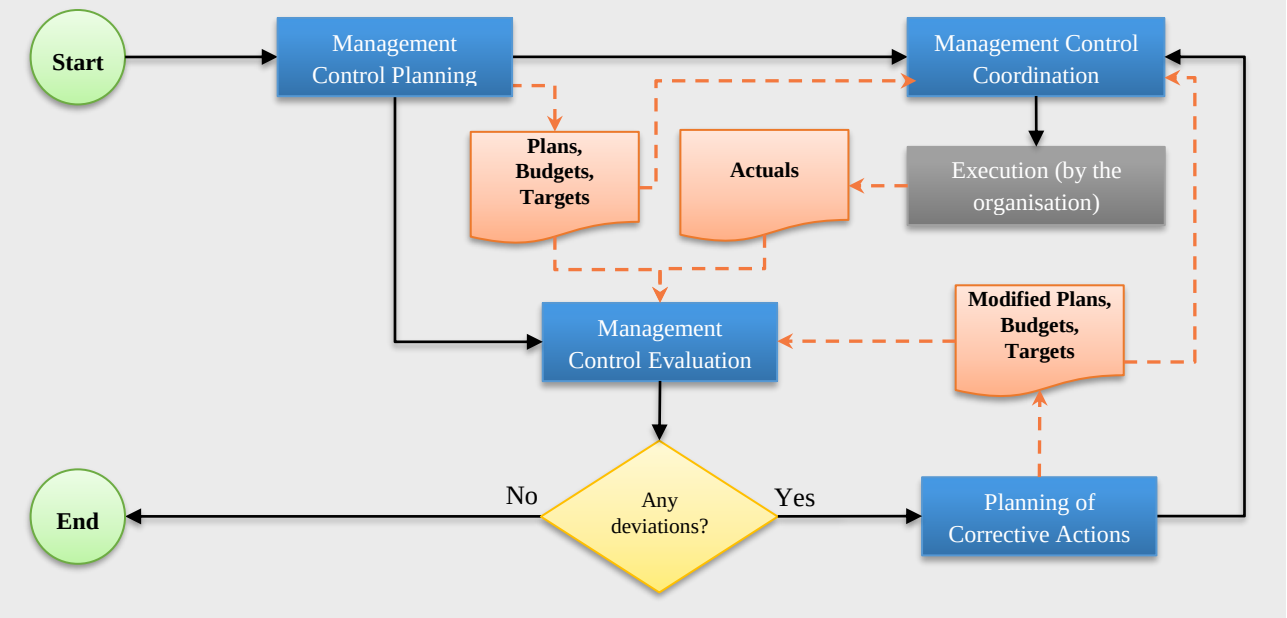
Target costing: One of the latest cost planning methods is the, so called, target costing. This method is a special, iterative version of the classic controlling planning method. The model is based on contribution margin calculations. In case the market price of the product or service offered by the organisation is less than the forecasted cost, then a series of cost optimisation and reduction procedures are applied[31] in several iterations to determine the cost level that provides targeted profit level. The method requires extensive, up-to-date information of processes, costs, customers and

[30] Using surveys like time-sheets for example
[31] For example, value and bottleneck analyses.

markets. Therefore, the efficient operation of a target costing system requires the integration of the information management and services function of the management control system.

2.4.3 Management control supervision in practice

Following planning the second control function, management control supervision, activates. The main objectives are: the constant monitoring and comparison of the actuals versus the budget; regular management reporting for both strategic and operative management and the coordination of the corrective actions. Figure 9 shows the practical process of management control supervision. This process is also called the small regulatory cycle[32] of controlling [KÖRMENDI AND TÓTH, 2006].



Source: Compiled by author based on figure 7 from [KÖRMENDI AND TÓTH, 2006, p.37.]
Figure 9: The practical process of management control supervision

The effective management control supervision should not only identify the deviations but also analyse the root cause to improve following planning cycles and to ensure that there is enough time for corrective actions. Management control evaluation and supervision is split by the following systems:

Operative supervision: operative supervision in practice is the computer based analysis of the deviation function. The analysis can only be carried out effectively if the adequate level of information is available. Therefore, the collected data should be granular enough to cover cost centres and products (cost objects) across the entire organisation in harmony with the organisational structure, strategy and operative plans. The data should be collected on a transactional level but at least by daily aggregates. The budget versus actual gap analysis should follow the same level of granularity.

Strategic supervision: Strategic supervision is carried out on a less frequent periodic basis (quarterly, half yearly, annual, etc.). It is the aggregation of the results of the operative supervision on the one hand. On the other hand, other strategic, project and action related evaluations are carried out as part of the strategic supervision.

[32] The large regulatory cycle is the feed-back loop between the operative and strategic operative sub-system, which works similarly to the small cycle whilst the large cycle prepares corrective strategic and tactical actions instead.

The results of the comparison of the actuals versus budget and the root-cause analysis are then reported to the different levels of management (ranging from team leaders through senior managers to executives). The management control supervision also plays a key part in the identification of alternatives and corrective actions (using scenario analysis and sensitivity analysis for example); decision support; and in the coordination of execution. In case of a constant negative trend, management control highlights the risk. However, decisions are escalated to senior management.

In the last part of Chapter 2.4 author continues to narrow down the knowledge gap identification by reviewing the marketing management control specific planning and supervision.

2.4.4 Marketing budget planning and supervision

Marketing budget planning determines the total annual expenses of all intended marketing activities (marketing communications and advertisement etc.). This planning is part of the above detailed management control planning procedures. Typically used in a bottom-up workflow where the marketing specific control area determines the initial area specific budget based on the strategy and plans for the year. The practice and literature of marketing and advertising budgeting concentrate on methods *"ranging from arbitrary "rules of thumb" to complex management science models"*. [PIERCY, 1987, p. 45.] Current dissertation recognises two approaches to advertisement budgeting: the heuristic and the analytical approaches. [REINECKE AND FUCHS, 2006].

Heuristic budgeting approach: Heuristic budgeting methods are primarily based on the planning practice and experience of previous planning cycles as well as estimated activities. These approaches are simple however are limited as they focus on one main driver in general.

- **Budget based on the budget of the previous period:** The main advantage of this method is the simplicity however does not consider the changes of the environment or the customers.
- **Budget based on revenue or profit:** Based on empirical, historical or benchmark data, the management determines the percentage value of the economic year's profit or revenue that the company will spend on marketing in the following year. Although the method is simple it does not promote efficiency or development of the marketing methods. [STRAUSS, 2008]
- **Budget based on available financial means:** Similarly to the previous approach, a percentage value is determined by the management but in this method the basis of the calculation is the planned (next year's) profit. This method reflects the growth aspiration of the company however inflexible in crisis situations.
- **Budget based on competitive activities:** This marketing budget is determined based on a series of competitor benchmarking activities. The budget for marketing is based on the companies *"share of voice"* among competitors. [DILLER, 1998] This method is effective in a competitive markets however inefficient use can lead to over-budgeting.
- **Budget based on communication objectives:** This is a bottom-up approach calculated based on the predetermined planning objectives and related communication activities. [RIEG, 2015] This method is potentially more accurate than the previous ones however requires the constant review of marketing activities and costs as well as the re-evaluation of marketing tactics.

Analytical budgeting approach: Existing analytical approaches mainly focus on the correlation between the communication budget and a target value applying basic statistical procedures. Based on the target value we can differentiate the following analytical approaches [STRAUSS, 2008]:

- **Potential-related:** This model aims a certain product or brand awareness and promoter scores
- **Market success-related:** Market success-related models target the relationship of marketing spend with the number of quotes, sales, leads or referrals. The underlying assumption is that there is a positive correlation between the variables.
- **Dynamic budgeting approach:** This approach allows management control professionals to understand the system of multiple variables (goodwill, wear-out effects[33], carry-over effect[34], forget rate, reactance) at the same time however this requires advanced statistical methods and increased amount of data compared to the previous methods. [RIEG, 2015]

Although the heuristic methods are preferred for the simple, less data intensive approach, current dissertation is intending to extend the analytical approach with another, statistics and data science based, predictive method, which with machine learning, can offer a sophisticated and automated approach in the future.

Marketing Budget (spend) Supervision has two main functions: marketing evaluation and marketing audit. Historically marketing audit was considered as the key strategic audit function to check marketing performance, and improve efficiency [REINECKE, 2007] but with the rapid development of the management control science the two functions have been separated and certain tasks and objectives of marketing audit became part of the supervision function of marketing management control [KÖHLER, 1993].

The **Marketing evaluation** literature [KÖHLER, 1993] differentiates two main activities:

- **Process control:** Process control is an element of the marketing (budget) evaluation. It is the constant monitoring and analysis of the marketing related activities, lead times and deadlines. Often the key stages of the sales process (e.g., the activities of advisors, agents or sales staff, etc.) are monitored and reported to manage performance and identify deviations from the plan.
- **Result control** is the complex process of three specific control activities: *Impact control* (it constantly measures the impact and the aggregated progress of marketing objectives, it answers the question: "Are we doing the right things?"); *efficiency control* (it is the assessment of the efficiency of marketing activities, identifies bottlenecks and other activity related issues and answers the question "Are we doing things right?" [DRUCKER, 1993]); and *cost control* (it is the comparison of the actual versus planned costs to identify budget constraints or the need for budget rearrangement).

Marketing audit: marketing audit (in opposite to marketing evaluation) assesses the operations, processes strategy, environment and the tools of marketing management instead of the impact of marketing decisions. [HAJDÚ, 2013]; [KOTLER AND KELLER, 2012] While marketing audit is a holistic periodic review of the marketing management system, marketing management supervision is a continuous review of the efficiency of processes and the management of objectives.

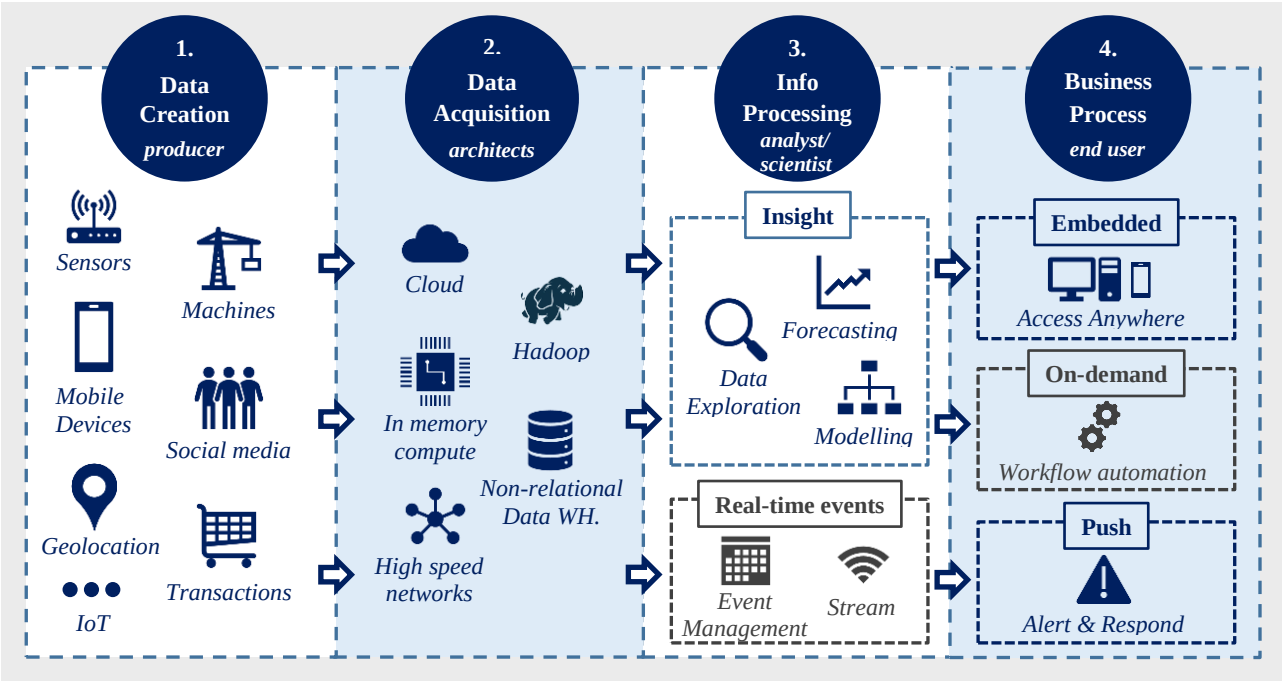
[33] The impact of advertisement is decreasing over time due to the aging of the advertising effect.
[34] The impact of advertisement is delayed until a later time period.

2.5 ENABLERS OF IMPROVEMENT: BIG DATA & DATA SCIENCE

So far we have reviewed the evolution of the management control theory, the different functions and objectives as well as the structure of management control systems at the organisation. We have also revised the specifications of bank and marketing controlling where the increased importance of information technology and the information management services function became evident. During the assessment of the management control planning and evaluation techniques, author identified an opportunity to introduce a new, applied statistics based cost planning and evaluation method to the existing range of processes. In the current chapter author is aiming to highlight the rapid development of the, so called, 'Big Data' technology and Data Science, which makes cost prediction and forecasting, in a traditionally accountancy driven planning area, a viable and efficient method.

2.5.1 Big Data and Data Science

In many industries 'Big Data' represents opportunities and challenges for the organisation. The term actually refers *"to the vast amount of data continually collected through devices and technologies such as credit cards and customer loyalty cards, the internet and social media and, increasingly, WiFi sensors and electronic tags. Much of this data is 'unstructured' – data that does not conform to a specific, pre-defined data model."* [CHUA, 2013] It is estimated that by 2020 there will be more than 16 zettabytes[35] of useful data captured and stored around the globe. [TURNER et al., 2014] Big Data terminology also refers to the innovative technologies responsible for the collection, storage and distribution of data. Figure 10 shows the Big Data Ecosystem from data creation to decision. The ecosystem covers four main areas:



Source: Compiled by author based (IoT refers to the Internet of Things) based on [STIRES, 2013]

Figure 10: Big Data Ecosystem - From data to decisions

1. **Data Creation:** In the last decade, due to the rapid development of multiple technologies, including wireless communication, real-time analytics, machine learning, commodity sensors,

[35] 1 zettabyte (ZB) = 1 trillion gigabytes (GB) = 10^{21} bytes (B)

and embedded systems or even contactless payments, faster and real-time transfers [KOVÁCS AND PÁL, 2012], etc., more and more devices, machines and other everyday devices started to be connected to the internet, which generates very granular level data from a wide range of events and activities around the world. [MCEWEN AND CASSIMALLY, 2014]

2. **Data Acquisition:** The second element of the Big Data ecosystem is the layer responsible for the collection, systematic storage and distribution of data. The converging development of the following three key elements enabled the wide spread of big data and data innovation: (1) The price decrease of RAM and processors enabled faster processing power, which (2) combined with a more efficient processing management algorithm and data storing system, such as Hadoop, provided the platform and infrastructure for large amount incoming data and outgoing information processing and (3) the maturity of high speed data transfer networks combined with the cloud technology enabled user to exchange data almost anywhere and anytime.
3. **Information Processing:** In order to efficiently use this wealth of data, traditional data management and extraction methods are not adequate anymore. The new data types, granularity and history of available data combined with the increased processing power enables real time, accurate data exploration, modelling and forecasting furthermore, using machine learning, fully automated event management (e.g. real-time fraud detection and prevention) is possible. This requires a collection of mathematical, statistical and programming methods combined in the, so called, Data Science, a new data-intensive approach to business utilisation and scientific discovery [HEY et al., 2009].
4. **Business Processes and application:** The final layer is the application of the results delivered by the first three layers. Businesses can monetise the newly acquired information without any manual interactions, furthermore decision makers can access contextual data anytime and anywhere. Users can be sent real time alerts and their responses can be automatically processed. Besides, traditional workflows can be fully automated just to mention a few of the possible business application.

On the one hand, if the organisation identifies the right way to harness the information it can improve performance, productivity and it also can create new income streams and competitive advantages. On the other hand, a poorly designed system can lead to breaches of data security and decrease productivity. Many new and existing organisations are now building their core activity around their ability to collect and analyse information in order to extract business knowledge and insights. The adoption of this technology and the preparation for creating value from the vast amount of information is not a luxury anymore, it is becoming a basic requirement for survival in the modern markets. [CAVANILLAS et al., 2016]

Data Science: As discussed above the new variety, volume and velocity of data requires a combination of new skills and methods in order to utilise the opportunity presented to businesses, researchers and analyst. Data science, or data-driven science, is an interdisciplinary field of scientific methods (mathematical and statistical dominantly), processes (programming) and systems to extract knowledge or insights from data in various forms, either structured or unstructured. [DHAR, 2013] Data science has several dimensions: exploratory data analysis; information design; interactive data visualization; descriptive statistics; inferential statistics; statistical graphics; data analysis; infographics etc. Data scientists using a range of skills clean, process and present the data to uncover hidden solutions to business challenges therefore it is essential to have technical (mathematics,

statistics, SQL/python/C/Java programming, data mining, data visualisation, Hadoop etc.) and business skills (analytical, good communicator, industry knowledge, etc.) at the same time.

2.5.2 Data science in the service of management control

When the powerful Big Data and Data Science meet the information and future oriented management control concept the number of practical applications and improvements of existing management control systems is almost endless. Table 7 is the summary of opportunities and challenges big data presents to accountancy, management control and finance professionals.

Table 7: Opportunities and challenges big data presents to controlling

Area	Opportunities	Challenges
Valuation of data assets	• Helping companies value their data assets through the development of robust valuation methodologies • Increasing the value of data through stewardship and quality control	• Big data can quickly 'decay' in value as new data becomes available • The value of data varies according to its use • Uncertainty about future developments in regulation, global governance and privacy rights and what they might mean for data value
Use of big data in decision making	• Using big data to offer more specialised decision-making support in real time • Increase collaboration with other departments to determine when big data can most usefully be shared with internal and external stakeholders	• Self-service and automation could reduce the need for standard internal reporting • Cultural barriers might obstruct data sharing between silos and across organisational boundaries
Use of big data in risk management	• Expanding data sources used in risk forecasting to see the bigger picture • Identifying risks in real-time for fraud detection and forensic accounting • Using predictive analytics to test the risk of longer-term investment opportunities in new markets.	• Ensuring that correlation is not confused with causation when using diverse data sources and big data analytics to identify risks • Predictive analytic techniques will mean changes to budgeting and ROI calculations • Finding ways to factor failure-based learning from rapid experimentation techniques into processes, budgets and capital allocation

Source: Table 3.1 from [CHUA, 2013, p.14.]

As we can see the combination of the scientific areas presents opportunities. Management control systems by 2020 will be able to create accurate insight almost real-time; identify innovation through new business models; improve the quality of planning and forecasting; and contribute to business-critical decisions about strategy and investment based on a wider range of data. Management control professionals who adapt to the new challenges by upgrading their existing knowledge and learning new skills in big data and analytics will be able to help organisations to: improve planning and evaluation; improve decision making; identify and develop new markets; improve operating efficiency, which together ultimately leads to increased profit. [CHUA, 2013]

As we have seen in the literature, the paste of change is increasing, with the latest development in the fields of Big Data and Data science the volume, variety and velocity of information is rapidly increasing which provides a firm foundation for improvements on other data and information driven scientific and management areas such as Management Control. The literature review has also highlighted that heuristic marketing budget and cost planning and evaluation methods now can be extended efficiently with other statistics and data science based, predictive methods. Based on the literature review the hypotheses are defined in the next chapter.

"The greatest value of a picture is when it forces us to notice what we never expected to see."

John Wilder Tukey (1915 – 2000)

3. MATERIAL AND METHOD

In this chapter author is presenting the hypotheses of the research based on the research strategy presented in Chapter 1 and the literature review in Chapter 2. Current chapter also contains results for one of the key aims, namely to set up a methodology to support financial analysts and controllers to conduct marketing campaign cost prediction and evaluation by describing the unique sequence (process) of statistical methodologies required by analysts to complete the cost prediction exercise.

3.1 RESEARCH STRATEGY AND HYPOTHESIS

In accordance with the empirical observations, the research aims, strategy and the comprehensive review of the relevant literature, author aims to test the following hypotheses:

Hypothesis 1 (H1)
The existing cost planning, especially marketing campaign cost planning, methods are dominantly non-statistics based approaches. Most (over 75%) of the applied planning (especially cost planning) and evaluation methods described in the selected literature are non-statistical approaches (do not apply multivariate statistical methods for example).

Hypothesis 2 (H2)
The field of management control systems harnessing big data technology and data science is new and has only been emerging from the literature in the last decade. The literature of management control harnessing big data technology and data science is still developing (it shows an upwards trend with no evidence of a plateau yet).

Hypothesis 3 (H3)

Hypothesis 3.1 (H3.1)
In order to provide a statistic based and potentially more accurate and efficient campaign cost planning and evaluation method, a mathematical model between the campaign goals, the marketing costs and other contributing factors must exist. *A mathematical function (model) exists between key bank marketing goals (such as the number of new customers acquired or recruited or market share per campaign) and campaign costs with other factors.*

Hypothesis 3.2 (H3.2)
If such model exists then the start-to-end process to support analysts and controllers to conduct marketing campaign cost prediction and evaluation analysis should be captured. *The statistical process, from the conceptualisation of the modelling aim to the statistically validated regression equation, can be captured on simple process diagrams to support financial professionals.*

Hypothesis 4 (H4)

Hypothesis 4.1 (H4.1)
There is a linear relationship between the campaign goals, the campaign spend and most of the other contributing factors (therefore multivariate linear regression is an adequate method).

Hypothesis 4.2 (H4.2)
Using the modelling database and the multivariate linear model, the key factors that determine success of a bank marketing campaign (such as number of new customers and size of market share per campaign) can be identified with their weight (or contribution) to the success.

Hypothesis 5 (H5)

Hypothesis 5.1 (H5.1)
The linear equation extracted with the multivariate linear model can be used for cost forecasting and identifying required cost levels for new customer recruitment targets in the system of other contributing factors for future campaigns. *The cost values forecasted by the model for new observations are statistically significant (with other words, the correlation between the forecasted and actual spend values is over 65% and the variances of the forecasted and observed populations are equal at the 95% confidence level).*

Hypothesis 5.2 (H5.2)
The final assessment is the uplift in prediction accuracy. *The campaign cost values predicted by the multivariate linear model are closer to actual values compared to the campaign cost calculated with existing marketing control planning methods to the actual values.*

3.2 DATA SOURCES AND DATA COLLECTION

Author has structured the research and the methodology in a way that it serves two main purposes: 1.) the results of the hypothesis testing will provide scientific evidence to prove or reject the original research questions therefore it contributes to the academic knowledge; 2.) in case, the hypotheses are proven and the new methodology is statistically and practically valid then current research could be used by practical management control and financial professionals too almost like a manual guide.

3.2.1 Primary data sources

Author has conducted a primary qualitative research in 2016. The aim of the interview based research was to collect a wide range of practical factors that influence and potentially predict the change in bank marketing targets (new and existing customer recruitment for different product categories and new business market share). As Chapter 3.3 details, the new methodology outlined in the dissertation requires the collection of explanatory or predictor variables of a target variable defined by the objectives of the organisation in a preceding step. If the method is applied by a practical professional at a single organisation then the variable collection can be done as part of a less formal interview with internal management control professionals only. However, author for the current research aimed to involve professionals from multiple financial organisations in order to capture views from a wider range of bank experts. Qualitative research method was chosen to extract the complex logic of

management control specialists as this method is interactive and allows the researchers to uncover deeper content, which would not be possible with a survey.

Author has contacted 15 industry professionals of which 9 participated anonymously in the research. The financial organisations of the participating experts cover over 68% of the retail banking industry[36] in the United Kingdom. The organisations were selected in a way that their managers would have similar challenges as the financial institutions offer the same product range, have multiple distribution channels (online, telephony distribution and branch network) and are present in all parts of the United Kingdom (Scotland, Wales, England). Table 8 summarises the characteristics of the interviews.

Table 8: The characteristics of the sample and interviews

	Asset size[37]	Employees[37]	Role	Date	Location	Duration	Recorded
1.	Over £500bn	Over 50,000	Senior planning analyst	28/01/2016	Edinburgh	87 minutes	Yes
2.	Over £500bn	Over 50,000	Financial professional	04/02/2016	London	53 minutes	Yes
3.	Over £500bn	Over 50,000	Analytics manager	09/02/2016	London	47 minutes	Yes
4.	Over £500bn	Over 50,000	Financial Officer	03/02/2016	London	42 minutes	Yes
5.	£100bn-£500bn	10,000 – 49,999	Financial Controller	05/02/2016	London	61 minutes	Yes
6.	£100bn-£500bn	10,000 – 49,999	Senior Business Analyst	05/02/2016	London	56 minutes	Yes
7.	£100bn-£500bn	10,000 – 49,999	Head of Planning	10/02/2016	Leicester	39 minutes	Yes
8.	£0bn-£100bn	Up to 9,999	Senior Finance manager	27/01/2016	Newcastle	55 minutes	Yes
9.	£0bn-£100bn	Up to 9,999	Senior Finance analyst	25/02/2016	London	52 minutes	Yes

Source: Compiled by author

3.2.2 Secondary data sources

In order to complete the current research and potentially contribute to the academic community through the results, the data used for the model building exercise should reflect actual observations of the assessed environment. At the same time using and making publicly available commercially sensitive information is a risk that no privately held organisation can take. Author's solution for these contradicting interests is to extract the required data using SQL programming language from the databases of a participating retail bank (Bank A) and assess the interrelations between the variables. Then to modify (multiply or subtract) all figures with the same randomly generated number and assess the interrelations again. If the interrelations have not changed then to use this, simulation database. In order to ensure that not even the simulation database is accessible publicly, author's solution is to complete all data analysis and modelling using the simulation database on the financial organisation's own network and to use only processed and aggregated desensitised outputs in current research. This way the breach of data security can be avoided and a robust model and methodology can be set up and tested using actual data. For the model building, a mix of internal (*Accounting Database, Marketing Database, Business Intelligence Database, Management Information System with Market Research Agency data*) and external (*ONS, Bank of England, OECD, Metoffice, The Government of the United Kingdom and various newspaper articles*) data is used. The sources used in current dissertation are detailed at the relevant part of the methodology and results.

The following chapter serves two main purposes. On one hand, it provides a detailed description of the methodology applied to test the hypotheses defined above. On the other hand, in case H3 hypothesis is confirmed then the procedure notes and process diagrams confirm H4 hypothesis.

[36] Based on the percentage share of the total consolidated UK asset value at the end of December 2015.
[37] Source: http://newcityagenda.co.uk/wp-content/uploads/2014/11/Online-version.pdf

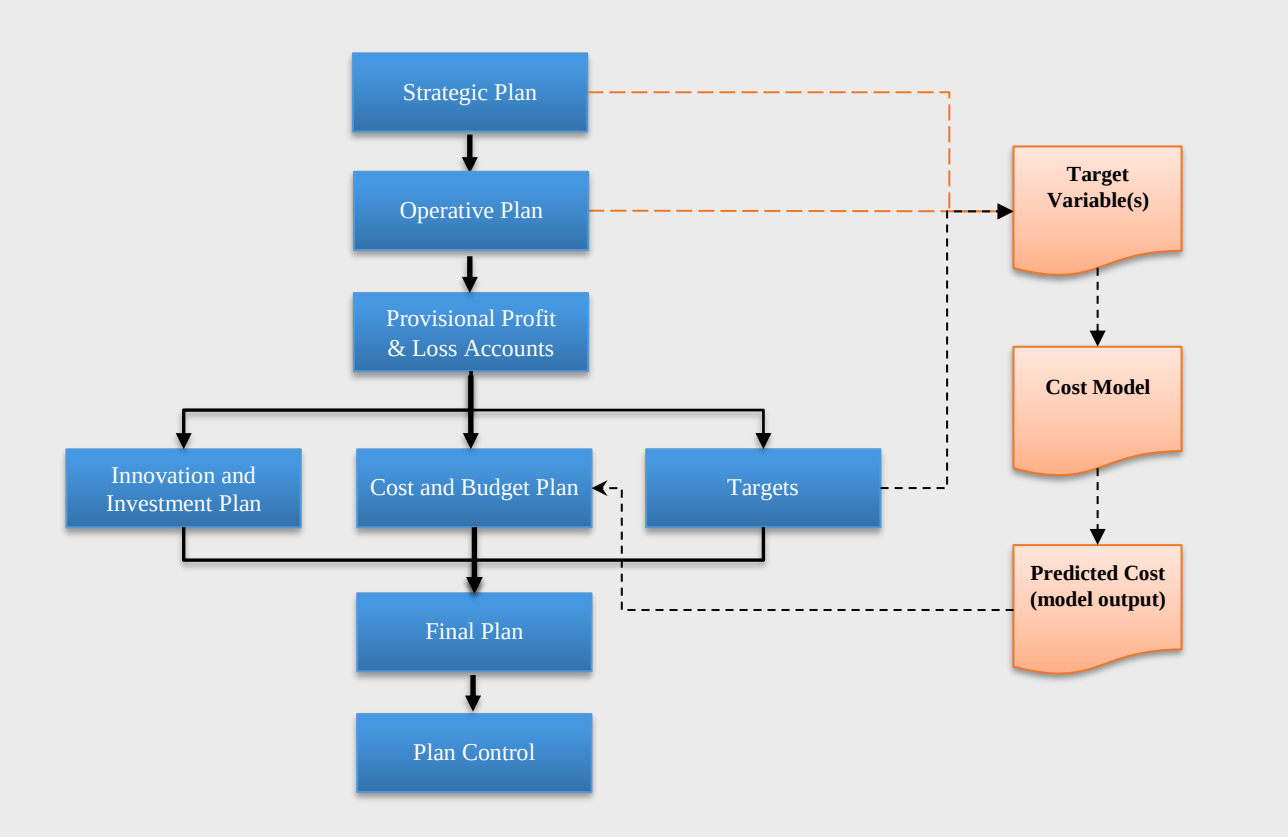
3.3 BUILDING A STATISTICAL EXPERIMENT: THE PREDICTIVE COST MODEL

In this section author is presenting the process of predictive cost modelling established by the author for the research. Although an analyst can have different reasons to build a regression model, current study is focusing on the predictive, control and inferential modelling aims.

3.3.1 Stage I. – Identification of the strategic aim

The first stage of the predictive cost modelling is the identification of the strategic aim. This strategic aim will then be translated to a strategic or target variable. Ultimately, the analyst will attempt to model the existing system with the aim of explaining the change in the value of the target variable through the complex network of explanatory variables.

Although predictive cost modelling is part of the operative planning process, the analyst always has to understand the relationship between the vision, mission, strategy and operative plans (as described in the Chapter 2.). Figure 11 shows that the analyst will identify the target variable(s) using the annual targets of the planning process. Then the output of the cost model will be the input to the cost and budget plan.



Source: Compiled by author based on Gyulainé-Jágerné, 2006 pp. 11.
Figure 11: The place of predictive cost modelling in the planning process

The first step therefore is to **understand the current strategy** of the financial institution. In order to increase the employee engagement, it is in the organisation's best interest to communicate the overall strategy and provide clarity on targets. [WELCH, 2011] Although the analyst should already be aware

of the corporate direction, the following sources can be used for modelling[38]: 1., Internal sources (strategy or tactical documents, operational budged plans, financial plans, product launch documentation, etc.); 2., Public documents (Investor Relations communications, annual statements, shareholder information packages, etc.); 3., External sources (studies, conferences papers, market research agencies such as CACI[39], eBenchmarkers[40], etc.).

Banks, similarly to any other businesses, make profit through their products (e.g. paid for products or margins/transfer margins) and services hence the aim is to increase the number of products purchased and services used by the customers. Financial institutions have 3 main ways to improve income: a) through the acquisition of new customers and the encouragement of existing customers to purchase additional products and use additional services; b) to improve the efficiency of existing product propositions and services. Due to the calculation capabilities of the new data science technologies (e.g. Hadoop, etc.) banks can have precise cost to serve and income calculations assigned to unique customers or products to help shaping strategy or identify opportunities for improvement; c) identify new income streams (e.g. big data monetisation, etc.).

In the current financial environment where customers have a wide range of financial providers to choose from with the transparency and speed of internet, it is ever increasingly difficult to recruit new customers or even to deepen the relationship with existing customers enough so the customers would fulfil all of their financial needs through one provider instead of shopping around for cheaper alternatives.[41] It has become vital for banks to evolve their strategies for competitiveness in the context of fast paced changes within technological, economic, social and regulatory environments. In spite of global challenges, the banking industry provides immense opportunities through expansion in new markets, technologies, and through personalisation to enhance customer relationship. Instead of focusing on financial targets only, financial institutions need to balance their strategy so they produce value for their shareholders, customers, staff and the society at the same time. With the explosion of digital technology, particularly customers' use of mobile phones and tablets, banks face new challenges with respect to shifting customer behaviour whilst customers still interact with banks through online sites, branches, ATMs and call centres which means financial institutions also need to connect to their customers through multiple channels. [KUMAR, 2014]

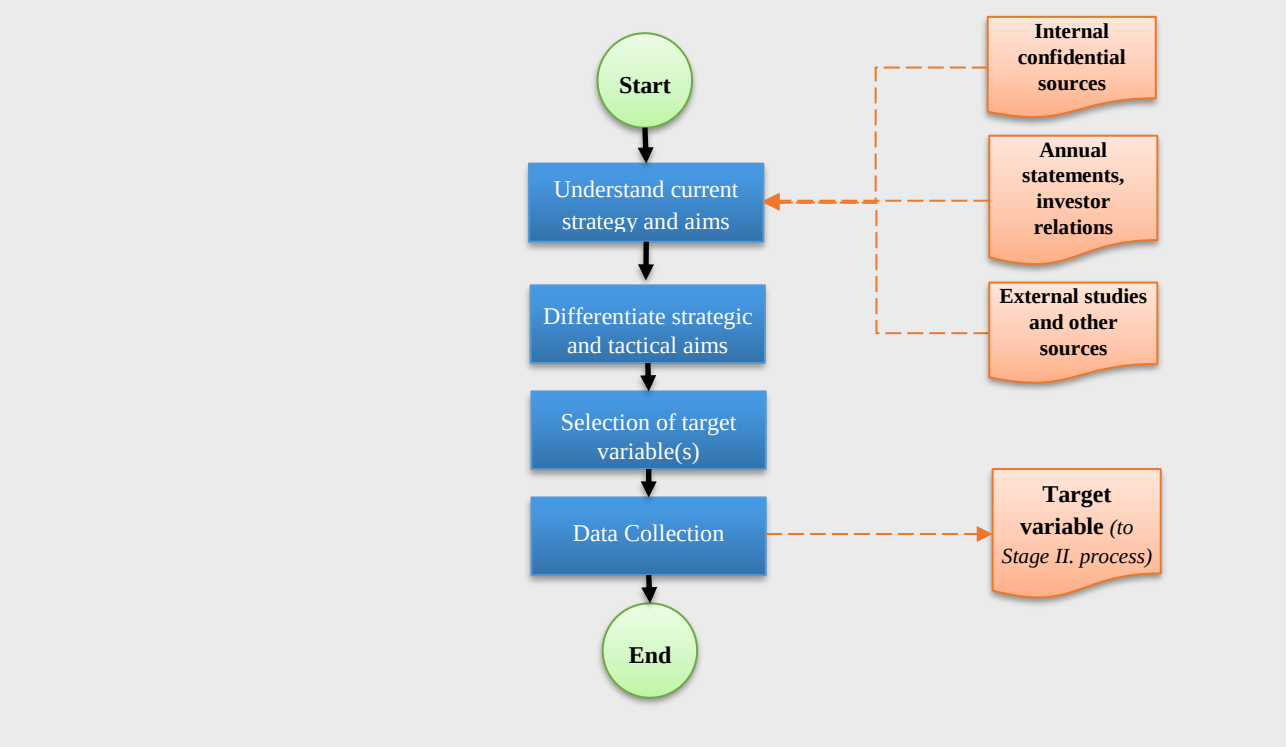
When the analyst is searching for a target variable it is likely to be linked to number of customers the bank would like to acquire, number of products to be sold or service to be used but more recently banks form strategies around the level of customer satisfaction, product or market share, customer behaviour (e.g. "aim for the bank is to reach 2 million active digital customers by 2020"), etc.

The second step is the **identification of the time horizon to differentiate between tactical and strategic aims**. Organisational strategy is covering a strategic time frame of 3-5 years. This strategy is normally broken down by tactical (1-2 years) aims and operative targets (within a 1 year). The differentiation is important as the analyst has to know whether the target variable needs to come from strategic aims or the cost model needs to be built for a specific campaign aim or target. For example in countries where the cost of funding is cheap (due to low base rate or government supported funding

[38] Please note that the list is not complete as there might be other sources available too.
[39] For more information please visit https://www.caci.co.uk/
[40] For more information please visit http://www.ebenchmarkers.co.uk/
[41] The Fintech industry represents an additional threat to classic financial activities. Fintech (Financial technology) is an economic industry composed of companies that use technology to make financial services more efficient. e.g. eToro, GoCardless, Nutmeg, RateSeller, Cabbage, etc.

programs, etc.) the overall strategy of the bank is to increase the number of customers in strategic terms but to keep deposits low on tactical terms as interest paid on these deposit can actually be expensive for banks. In this case the analyst is looking to build the model around number of customers rather than market share or deposit amount. In other markets, like credit card or mortgages both the market share and the customer/contract volumes can form part of the group of strategic variables.

The last step is the **selection of the most appropriate target variable** followed by the data collection. Once most of the variables are selected and signed off by senior stakeholders a granular data history of the last 4-5 years is required to be collected (more information about data collection principles are discussed in the Stage II.). Only when this is available can the analyst process to Stage II.



Source: Compiled by author

Figure 12: Process diagram of Stage I – Identification of the strategic aim

Generally, it is a good practice to identify and prepare more than one variable as the analysis of different cost models can give greater understanding of the underlying processes and help the planning and evaluation processes.

3.3.2 Stage II. – Identification of the predictor variables

The second stage of the predictive cost model is the identification of the contributing variables. Once the strategic aim, with other words the target variable, has been confirmed by the senior management the analyst is required to identify the variables, independent internal and external factors, which will determine the value of the target or dependent variable. In theory, for the best result the analyst could be required to identify every single factor that has a contribution, even if it only means a marginal gain in the explanatory power of the overall model. Hypothetically, the aim of the analyst at Stage II. should be to build a model, which can describe 100% of the change in the value of the depended variable. In real life, such models can only exist in laboratory or simple, closed systems therefore the

first thing to be determined by the analyst is the amount of time could be spent on the cost modelling exercise. For instance, if the company has limited resources, which ultimately can be expressed in the number of working hours available, then it will determine the time spent on the different stages. Although the optimal time spend on each step is out of scope of current research, based on analytical experience a trained analyst would probably spend 30-35% of the total time spent on the cost prediction for the initial model build. Once the resource capacity is known the analyst has multiple ways to determine which variables are required in the model. The initial aims are the following:

1. **Identify the most** (within the available time) **possible theoretical explanatory variables** that has an impact on the dependant variable. As this step is only the enumeration of the variables it is important to emphasise that all potential logical variables are needed to be recorded regardless data availability and potential multicollinearity with other variables. Depending on the complexity of the modelling issue and the required level of sophistication and accuracy, the analyst can decide to complete this step using only the relevant literature, past recorded learnings and (his/her) own experience or alternatively to set up a, so called 'brainstorming' session with internal experts and financial professional.[42] There are multiple techniques to record the results of the 'brainstorming' around the modelling questions. The author's personal preference for visual organisation is Mind Mapping or for a more sophisticated and insightful result the use of System Dynamics approach. Mind maps are useful for a simple and quick organisation of the main and sub-topics/categories related to the main research question. A recent study has shown that a typical mind map is rather small; around 31 nodes large and most mind maps (just over 60%) are edited only for one day. [BEEL-LANGER, 2011, p.81] "*System dynamics*[43] *is a computer-aided approach to policy analysis and design. It applies to dynamic problems arising in complex social, managerial, economic or ecological systems literally any dynamic systems characterised by interdependence, mutual interaction, information feedback and circular causality*." [BURTON-OBEL, 2013, p.239] It requires considerably more time to build but also has greater potential to understand how complex systems work and can answer certain cost modelling question without the need of using regression modelling.

2. **Reduce the number of variables based on data availability**. Theoretically, most if not all data related to the previously collected variables should be recorded and available to use for the cost modelling. On the one hand, in the age of big data and data lakes most of the company information is stored and managed in one single system. On the other hand, in reality there are limitations particularly when it comes to external data and business intelligence. Most models require some level of interactions with competitor information. This could be confidential information which means it is not possible to use or could be an estimate data through a 3rd party agency for a given compensation. Even if the data would be collectable by an analyst, business rationality should make the priority call and eventually some variables have to be dropped. The analyst should be conscious of the potential impact as the missing data can jeopardise certain models and could result the cost planning process to default back to the original practice.

3. **Set the time horizon and frequency required for the model** *(with other words set the sample size)*. Once all potential variables have been identified and passed the first stage screening of

[42] It is also recommended to add external experts/consultants if the company has a cost effective option for inclusion.
[43] More information about System Dynamics is available at http://www.systemdynamics.org/

data availability, the analyst has to evaluate the required minimum number of observations, which in case of a time series based cost prediction, is ultimately determined by the granularity of the time series data or with other words, determined by the time horizon of the analysis. *''As with any statistical analysis that is computed using sample data, the size of the sample (n) in large part determines the 'value' of the statistical results of a multiple regression analysis''* [GROSS, 1973, p.17] The analyst has multiple options to choose from. The analyst can predefine a value of the squared multiple correlation coefficient [ALGINA- OLEJNIK, 2000] or can try to minimise the shrinkage of the squared multiple correlation coefficient [PEDHAZUR-SCHMELKIN, 1991] when tring to determine the minimal sample size. Recent studies that examined the minimum number of observations required for an accurate[44] prediction have found a *"definite relationship […] between the squared multiple correlation coefficient and the minimum sample size."* [KNOFCZYNSKI-MUNDFROM, 2008, p.431] In their study authors found that as the squared multiple correlation coefficient decreased, the sample size grew at an increasing rate. Although it is rather difficult to find a 'universal rule' for the minimum number of observations, statistical sciences advise a minimum of 50 observations [SZŰCS, 2004, p.173] for linear regression analysis in social sciences. For simplicity Table 9 summarises the advised level of granularity for predictive cost modelling.

Table 9: Suggested Data granularity matrix

Modelling question	Recommendation	Number of observations
Available 5years or more data history on existing or similar product	Use monthly or weekly data[45]	60-500 data points
Available 3-5years data history on existing or similar product	Use weekly data	150-260 data points
Available 1-3years data history on existing or similar product	Use daily data for the relevant campaigns (min. 90 days per year)	270-540 data points
Available less than 1year data history on existing or similar product	Use daily data	180-360 data points
New product and there is no similar product	Predictive cost modelling is not viable	N/A

Source: Compiled by author

It is important to emphasise that different modelling questions will require different time horizons, which eventually will determine the minimum number of data points or observations, with other the words it will control the time scales. For example, if the prediction is required to be completed based on one season only, such as a Cross Tax Year period, as all other seasons

[44] The above cited research aimed to find sample regression models that predict similarly to population regression models. More precisely, what sample size is needed to ensure with a desired amount of accuracy, that the sample regression equation will perform similarly to the population regression equation? These minimum sample sizes were determined by conducting a series of Monte Carlo simulations. The study determined minimum sample sizes for a wide range of population correlation structures. Current dissertation considered the prediction accurate if the correlation coefficient between the predicted values and actual values is greater than or equal to 0.75.

[45] Author does not recommend using data history older than 5 years as this data may have lost its relevance due to the changes in economy and other circumstances.

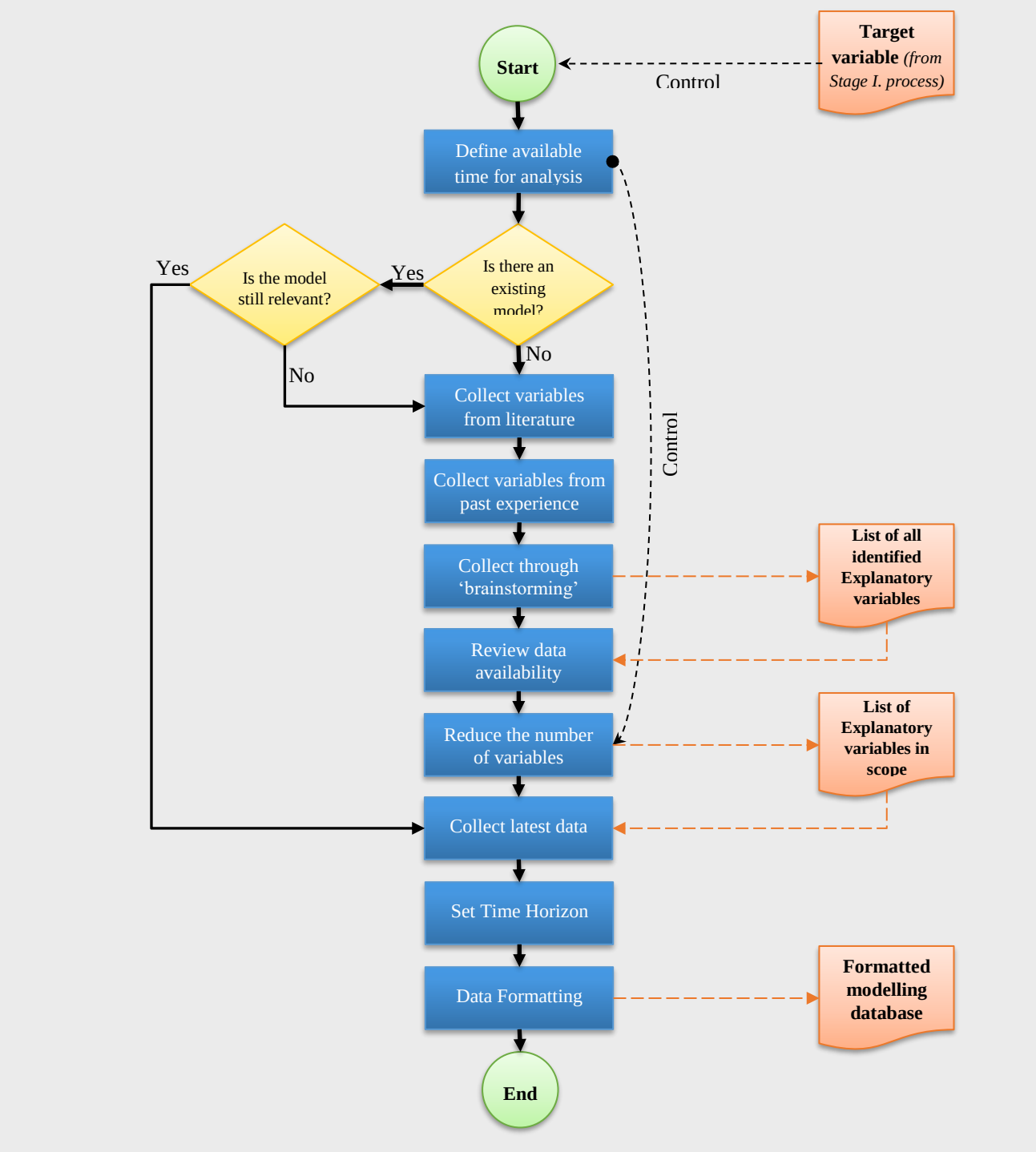
had significantly different circumstances[46] then the analyst is advised to split the data by day in order to have enough number of observations. At the same time, the cost predictive practice suggests[47] that weekly data of the last 3-5 years will provide the optimal number of data points that contain the relevant information and aggregation without the need of difficult daily data collection or daily phasing of a more aggregated data.

4. **Data preparation: data extraction, collection and formatting.** The last step of this stage is the preparation of the data. The execution of this step is vital and the actual data collection strategy should be chosen by the analyst based on the availability, type and source of the data. For cost modelling the analyst should collect the data by the chosen time scale where the dates are shown in consecutive rows. Each row means a data point uniquely distinguished by the date or time period it represents. The column of target values (the actual occurred cost values of each time period) should follow the previously described date column. All other (explanatory) variables should follow the target variable as columns aggregated appropriately for the time period. The term of 'appropriate aggregation' is used deliberately as the analyst has to determine whether the given variable requires a summation (e.g. sum of the volume of products sold), averaging (e.g. average of savings rate weighted by number of products sold by the different rates) or any other aggregation (e.g. minimum or maximum of house prices in a region within a month, etc.) whichever has the most explanatory power in the model. There are cases when the analyst can only use less granular data then the selected time scale. In this case, there are two options; 1) The analyst can decide to use a flat number for the date units covered by the variable (e.g. weighted average commercial savings rates are published only monthly by the Bank of England. Despite of our model is based on weekly data and we accept that the rates are constantly changing the change in the monthly rates should not be weekly phased as the phasing can reduce the explanatory power of the given variable as it is not accurate to assume no fluctuation in commercial savings rates within the month. In this case the same rate would be shown for all weeks within the month.); 2) In case it is accurate to assume a linear change or linear weekly trend then the analyst can create a weekly phasing by splitting the change into equal parts within the given time frame (e.g. GDP rates are also published monthly but the analyst can make a decision to split the increment between two consecutive data points and equally distribute it by the more granular time horizon). The analyst might also face another issue when trying to use partial data[48]. It is also possible for the analyst to include variables where full data history is not available in this case the suggested approach is to create a unique missing value instead of creating 0s as even 0 has a meaning within the model but the missing values can be excluded depending on the selected modelling approach.

Once all selected data have been collected and formatted the analyst can proceed to Stage III. – Pre-modelling analysis. Figure 13 summarises the end to end process of Stage II. The process flow highlights two key controls. The first one is the target variable, which is the output of Stage I. and

[46] Significance level in this case is determined by the analyst. Author would consider a change in the legislation system as significant difference for instance. The so called Payment Protection Insurance mis-selling row in 2006 resulted a significant change in the existing legislation. Using historic data after the changes would give irrelevant results as the future performance can no longer be predicted based on the past events. Please find more information on PPI mis-selling row here: https://www.theguardian.com/business/2011/may/05/how-ppi-scandal-unfolded.
[47] Based on the evaluation of 23 different cost predictive models.
[48] Data which wasn't available for the entire time covered by the predictive analysis but it has a measurable meaning on the target variable.

required for the process start. The second control is the definition of time periods for the analysis. These controls will have implications at later stages of the prediction process.



Source: Compiled by author

Figure 13: Process diagram of Stage II – Identification of explanatory variables

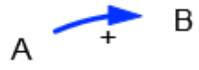
3.3.3 Stage III. – Pre-modelling analysis

The third stage of cost modelling is the pre-modelling analysis. Although experienced analysts often incorporate elements of this stage into Stage II and Stage IV, current study aims to highlight the importance of these steps and creates a separate stage. The main goal of the analyst at this stage is to

prepare the dataset for modelling with a statistical approach. As part of this stage the analyst can review the collection of data; apply an initial logical filter; transform the initial data to variables that serve the purpose of the model better and to understand whether there is a need for the reduction of the dimensions.

Assess the system of variables: The very first step of this stage is the assessment of the system of variables. The analyst should understand the logical relationship between variables. An approach, similar to the already mentioned system dynamics, is advised as it helps the analyst to explore the relationship between the variables and to apply an initial filter in case two or more variables would explain the same change. Although system dynamics is mostly used for the impact assessment of macroeconomic issues, political and business policies and has a flow approach [YEARWORTH, 2014] we still can adapt some of the logic and apply the same visual tools. The following basic notations [LANE, 2008] we can adapt from system dynamics are summarised in Table 10.

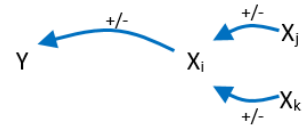
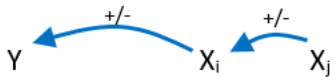
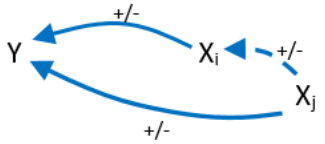
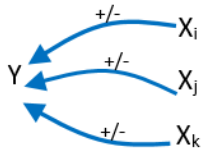
Table 10: Applied System Dynamics notations

Notation	Description
	The sign should be used when A variable causes B variable with positive link polarity. Ceteris Paribus[49] A increases then B increases and vice versa.
	The sign should be used when W variable causes Z variable with negative link polarity. Ceteris Paribus Z increases then W decreases and vice versa.

Source: Compiled by author based on LANE, 2008

The analyst should look for signs where two variables are in a relationship like where one depends on the other variable. In this case for example only one variable is allowed in the model. In order to facilitate this assessment, the analysts should build a simple graphical interpretation of how variables

Table 11: Basic relationship pattern

Scenario	Description and action	Example
	A variable is the aggregate of two or more variables. The suggestion is to exclude the aggregate variable as the components can give more insight.	One explanatory variable is the summation of other variables: $X_i = X_j + X_k$
	There is a functional relationship between two or more variables. The suggestion is to include the original predictor only.	One explanatory variable is the function of other variables: $X_i = aX_j$
	One of the explanatory variables explain directly both the target variable and another variable. The suggestion is to include all variables and carry out further tests.	The impact of the number of new customers and the number of internet banking logons on the product opening
	The variables are (seemingly) independent. The suggestion is to include all variables and carry out further tests.	The impact of weather, GDP and sporting events on purchase propensity

Source: Compiled by author

[49] All other things are unchanged.

are linked to the target variable and to see whether there are any (obvious[50]) interrelations between the explanatory variables.

This step should help the analyst to rationalise the number of variables to a certain extent however further analysis, based on objective methodologies, will help to make decision regarding ambiguous cases and identify other unwanted relationships in the modelling dataset.

Create artificial variables (dummy and proxy variables) where necessary: Analyst often face situations when there is a suspected relationship between the target variable and one or more qualitative factors. These factors are often not measured on ordinal scale furthermore sometimes not even captured as a numeric variable. There are also cases when a numeric variable would require further aggregation to capture the qualitative explanatory power. Depending on the number of outcomes of the variable the analyst can artificially create so called Bernoulli or dummy variables [DOMÁN, 2005]:

a) In case when the variable has only two values; either because it shows the existence of something (e.g. days in campaign vs. not in campaign and yes/no questions or simple flags and conditions, etc.); or because the actual variable has only two possible outcomes (e.g. gender: male/female or geographical identifier, such as city or town, etc.) the analyst can introduce a binary variable. (where for example 0 represents the non-fulfilled condition and 1 is the fulfilled condition or the existence to an answer). Variables with two potential outcomes can be treated in the same way but the analyst has to understand how to interpret the result in this case, e.g., the gender variable is often misinterpreted. There is no quality difference between any of the genders therefore using 1 and 2 to represent the two separate genders would potentially distort the analysis and make the interpretation difficult[51]. The only other option is to create the previously mentioned 0 and 1 numeric values but in this case whichever gender represents the value 1 will be the subject of the analysis, with other words if the analysis shows a week correlation between the dummy and the target variable it can mean two things at the same time; either the gender does not have a great impact on the target variable or only the specific gender that is represented by the value 1.[52] The analyst has to also be mindful of the multicollinearity when thinking about splitting the variable into two categories (such as Is the subject male? yes/no and Is the subject female? yes/no). During the planning and creation of the dummy variable the analyst has to understand what answer serves better the explanation and the hypothesis.

b) In case when the variable has more than two different values the analyst can create a, so called, proxy variable. The application logic and method is similar to the two-way dummy variables with the exceptions that in this case the analyst assign a value to the different values based on a hypothetical or existing hierarchy and normally 0 represents the non-existence of any of the elements. The most common application is around the assessment of the highest education. Although the analyst can choose to describe the level of education with the 'number of years spent in the education' it can be misleading[53]. A better option is to use categorical variable,

[50] For example a variable is the sum of two or more other variables, etc.
[51] Unless the analyst deliberately choose to grade the genders and set up a hierarchical meaning for the analysis.
[52] In order to isolate the two options, analyst often include other variables to test the other gender for example.
[53] The number of years spent in education does not necessary result higher education as someone can spend multiple years in the same year, etc.

which can eventually be translated to a numeric one. The analyst also has to keep in mind that the distance between two categories will impact the interpretation therefore it is important to predefine the hierarchy of value for the different categories. Continuing on the previous example the analyst has to decide whether a doctorate is as large step compared to a university degree as an A-level compared to a GCSE, etc.

Correlation analysis: The final step of Stage III is the initial analysis of the correlations between the target and explanatory variables. This helps the analyst to explore the interrelation of the different variables; to understand the nature and direction of the relationships and to start testing the initial model variables in order to avoid incorrect estimation (due to multicollinearity for example). In order to understand the relationship between the variables, a partial correlation matrix is needed to be calculated. It shows the strength and direction of the correlation between two variables after removing the effects of the remaining variables. The matrix is diagonally symmetric. The partial correlations for the diagonal elements are 1 (as the correlation is assessed between the same variable in the horizontal and vertical intersection of the matrix). The computation is simple:

1) Compute the standard correlation matrix;
2) Invert this correlation matrix;

$$R = \begin{bmatrix} 1 & r_{yx_1} & \dots & r_{yx_i} \\ r_{x_1y} & 1 & \dots & r_{x_1x_i} \\ \dots & \dots & \dots & \dots \\ 1 & r_{x_ix_1} & \dots & 1 \end{bmatrix}$$

where, R is the correlation coefficient;
y is the measured or dependent variable;
$x_1, x_2, \dots, x_i$ are the explanatory, independent or often called as predictor variables;

3) Compute

$$r_{ij} = \frac{-r_{ij}}{\sqrt{r_{ii}r_{jj}}}$$

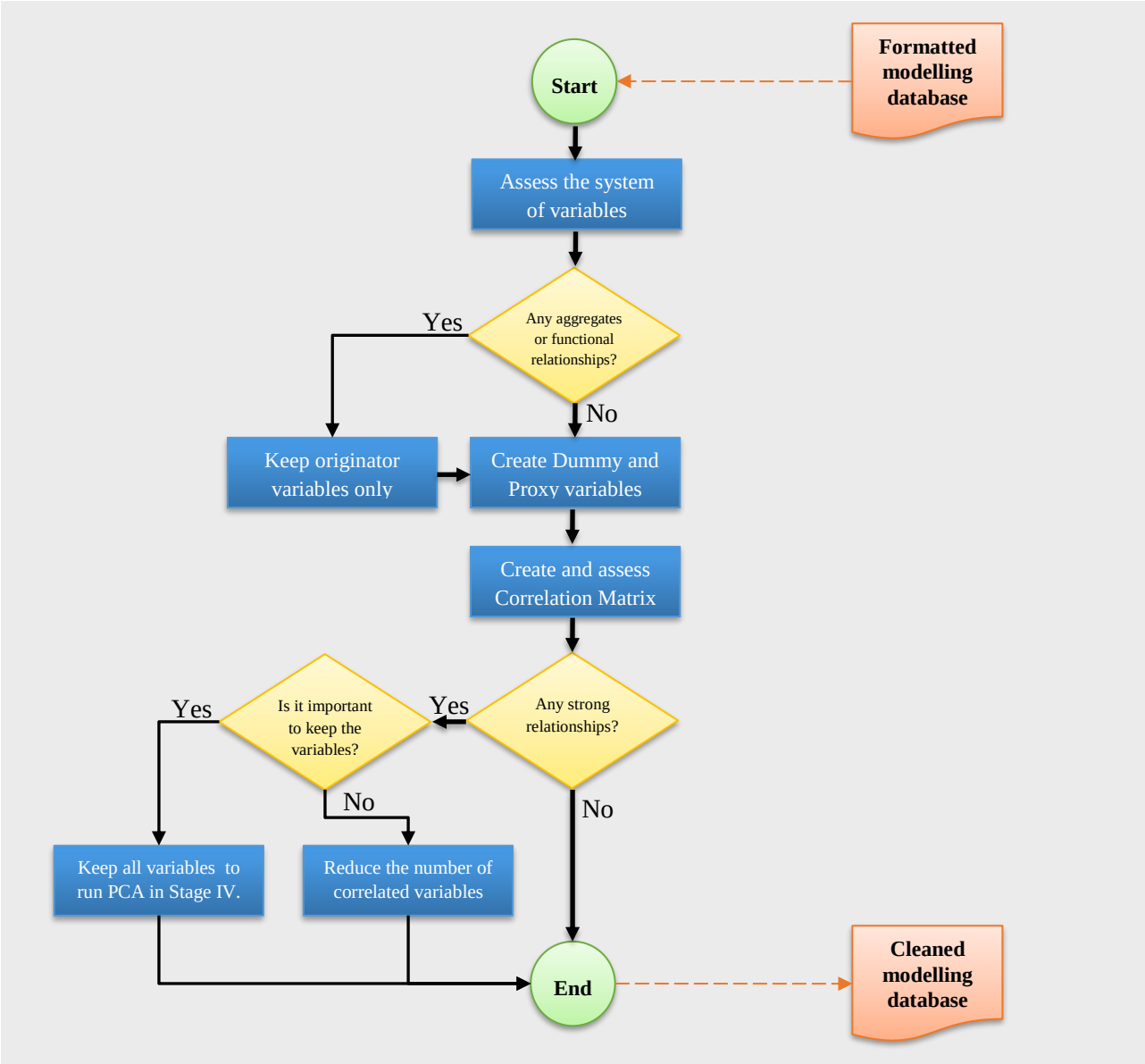
where r_{ij} is the (i,j)-th element of the inverted correlation matrix.

The analyst should review the calculated correlation matrix and identify variable pairs where the correlation is considered as strong. For reference: current research uses the following classification for the strength of correlation [SZŰCS, 2004, p.315]

- If the absolute value of correlation coefficient (|r|) is less than 0.25 then there is no relationship between the variables;
- If $0.25 \leq |r| < 0.5$ then there is a moderate relationship between the variables;
- If $0.5 \leq |r| < 0.75$ then there is medium relationship between the variables;
- If $0.75 \leq |r| < 0.1$ then there is a strong relationship between the variables;
- If $|r| = 1$ then the relationship is functional between the variables.

The analyst at this step can statistically test the variable pairs identified as potentially related. Once the variable pairs (groups) with medium to strong correlation have been identified the analyst has to make a decision whether one or the other variables could be excluded or the explanatory power should be transferred to the model. In this case the analyst should proceed to Stage IV in order reduce the

dimensions by compressing the meaning of the correlated variables into factors. The following figure summarises the suggested steps and decisions needed to be considered by the analyst.



Source: Compiled by author

Figure 14: Process diagram of Stage III – Pre-modelling analysis

3.3.4 Stage IV. – Reduction of dimensions

The next stage of the cost modelling is the reduction of dimensions, with other words the reduction of predictor variables. This serves two main purposes; firstly, it helps to eliminate potential weak to moderate multicollinearity[54] as correlating predictors with a latent factor are combined into one or more variables. Secondly, this process increases the accuracy of forecasting as the number of variables will not have a major impact on the parameter estimates. Furthermore, it makes the model

[54] The phenomena called multicollinearity is discussed in Chapter 3.3.5 under the 'Managing multicollinearity' section. Please note that Factor and Principal Component analyses are not suitable for high multicollinearity where $|r| \geq 0.9$ between two variables. This case requires theoretical reconsideration and potential elimination of one of the variables. [YONG-PEARCE, 2013, p. 88]

interpretation simpler. Statistical sciences offer a range of multivariate techniques for the analysis and simplification of variable interrelations as phenomena in natural and social sciences can only be described by the complex network of factors. The ultimate aim of multivariate analytical methods, such as Factor and Principal Component analysis, is to explore and describe the system of interrelations of these factors with a systematic approach. [SZŰCS, 2004 p.409]

Common Factor Analysis (CFA) and Principal Component Analysis (PCA) are statistical methods also used for the reduction of a set of observed variables into a smaller number of latent variables. The underlying assumption behind these techniques is a latent variable structure where the values of the groups of variables are determined by unobserved latent variables. These unobserved variables can be identified by the assessment of the partial correlations between the observed variables. [SAJTOS-MITEV, 2007 p. 283] The variables linked to the identified latent variables then can be replaced by the latent ones, which eliminates the harmful multicollinearity.

The first step of the variable reduction stage **is the method selection**, with other words deciding whether factor analysis or principal component analysis is required. Although factor and principal component analysis have similar methodologies and ultimately very similar purpose (data reduction, exploration of variable interrelations), the underlying assumptions behind these methods are different. As Table 12 shows there are some key differences between the approaches of the two methods:

Table 12: Similarities and differences between CFA and PCA methodologies

Metric	Principal Component Analysis	Common Factor Analysis
Observed variables	Assumes Linear relationship between observed variables with normal distribution for each observed variable. Each pair of observed variables has a bivariate normal distribution.	
Communalities	If communalities are large (close to 1.00) then the results of PCA and EFA could be similar	
Variance	PCA retained account for a **maximal amount of variance** of observed variables	EFA account for **common variance** in the data
Correlation matrix	Analysis decomposes correlation matrix	Analysis decomposes adjusted correlation matrix
Unobserved (latent) variables	Minimises sum of squared perpendicular distance to the component axis	Estimates factors which influence responses on observed variables
	Component scores are a linear combination of observed variables weighted by eigenvectors	Observed variables are linear combinations of the underlying and unique factors

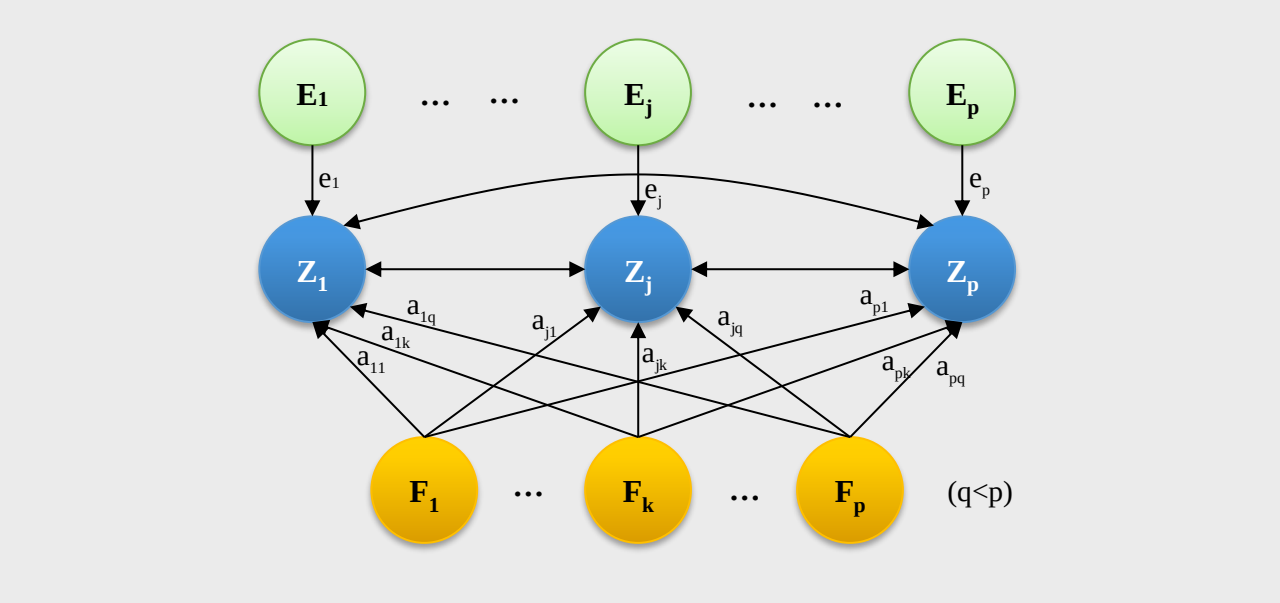
Source: based on SUHR, 2005 p.3

When the analyst is selecting the appropriate method the main consideration is the purpose of the reduction. In practice it means that:

- If the intention of the analyst is to explore the interrelations of the observed variable set, with other words to identify if there is a latent factor structure behind the observed variables then the common factor analysis is required to extract the underlying dimensions that account for the intercorrelations between the observed variables.
- If the aim is to produce the minimum number of composite variables or principal components accounted for the maximum part of the variance from the original set of variables, with other

words to compress the meaning of a selection of correlated variables with the minimal loss of information then the principal component method is advised. [55] [DSSC, 1995]

In cost modelling as **the second step the analyst can conduct an exploratory factor analysis** to understand if there are variables in the planned model that have a common variance driven by latent factors. Figure 15 demonstrates the relationship between the observed and the unobserved variables in the factor analysis.



Source: Compiled by author based on SZŰCS, 2004 p.416

Figure 15: Graphical model of factor analysis

The linear model of factor analysis can be written as follows:

$$Z_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1k}F_k + a_{1q}F_q + e_1E_1$$

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jk}F_k + a_{jq}F_q + e_jE_j$$

$$Z_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pk}F_k + a_{pq}F_q + e_pE_p$$

where, Z_j is the j^{th} standardised observed variable (X_j); a_{jk} is the coefficient that signifies the importance of the k^{th} factor variable and its weight affecting the change in the value of the j^{th} standardised observational variable; F represents the unobserved factors, E_j is the j^{th} specific effect and e_j is the relevant loading of it, also called the residuals.[56]

The following steps are required by the analyst to complete the factor analysis:

1. **Assessment of the prerequisites**: Before starting any analysis it is important for the analyst to set up the aims and establish the theoretical concept behind the common factor analysis, with other words the analyst has to understand the data, the available variables, the context

[55] Another consideration for method selection is the variance. In practice if the observed variables are relatively error free due to measure captured by the variable (e.g. years in business, etc.) then PCA method is advised unlike when the variables are only indicators (e.g. customer satisfaction) and therefore the error variance accounts for a significant portion of the total variance. In the latter case factor analysis is required.

[56] For more details on the mathematical algorithm behind Factor analysis please read SZŰCS, I. (2004): Alkalmazott Statisztika pp. 447-494.

and using this knowledge (he/she) has to establish assumptions or hypotheses specific to CFA. Only this way the analyst can achieve an effective variable reduction. The next step is the statistical assessment of the prerequisites and the existence of the underlying assumptions of CFA [SAJTOS-MITEV, 2007]:

a. The first step is to ensure that all variables included into the factor analysis are metric.
b. Then the analyst should assess the presence of correlation between the variables. The (previously detailed[57]) correlation matrix is generally used for this exercise. In this step the analyst is visually looking for signs of medium to strong correlation[58] between variables. It can mark a complex multi-factor latent structure if there are groups of variables with correlation instead of all or only some variables show connections.
c. The correlation matrix is followed by the visual assessment of the correlated variables. The analyst should ensure that there is a linear relationship between the variables in question by using scatterplots. At this stage a visual determination is sufficient.
d. The visual assessment of the correlation and the variance is followed by the Bartlett test[59], which helps to test statistically if the off-diagonal elements of the correlation matrix are significantly different to zero, with other words whether there is a correlation between the variables or the relationship is only due to randomness.
e. The next step is the analysis of the anti-image matrix. It helps to determine the variables we should exclude from the factor analysis based on the Measure of Sampling Adequacy (MSA). Generally, variables with values below 0.5 should be excluded and the process should be completed again. The anti-image matrix also helps the analyst to visually assess if the unique part of the variance (off-diagonal elements) is low enough[60] to continue with the assumption that the variance of the correlating variables is driven by an underlying latent factor structure.
f. The final step is the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy, which helps to determine if the sample is suitable for factor analysis. Similarly to the previous MSA assessment the dataset is adequate for factor analysis if the Overall MSA is above 0.5. [YONG-PEARCE, 2013, p. 88] However, it is advised to increase the level to 0.7[61] for a more accurate analysis.

2. **Initial extraction and the determination of the number of factors to retain**: Once the dataset fit for factor analysis, the analyst has to complete the initial extraction using the chosen statistical software and has to make a decision about the number of factors to retain using the following methods: Kaiser-Guttman criterion, Scree test or the proportion of the variance for each component keeping the interpretability of the model in mind. All three methods focus on the eigenvalues, with other words on the common variance of the variables explained by the factors. When determining the number of factors the analyst can:

[57] For details please see the Correlation Matrix section of Stage III – Pre-modelling analysis chapter.
[58] Where $|r| \geq 0.5$
[59] For details please see the Tests for constant error variance section of Stage V. – Multivariate Linear regression modelling chapter.
[60] Values between -0.7 and +0.7 are considered low for factor analysis.
[61] Kaiser Criterion is said to be reliable when: a) the averaged extracted communalities is at least more than 0.70 and when there are less than 30 variables, or b) the averaged extracted communalities is equal or above 0.70 and the sample size is above 250 cases. [FIELD, 2009, p.640]

a. Retain factors based on the eigenvalue. The Kaiser-Guttman criterion sets the cut-off value 1 which means the analyst can retain the factors with an eigenvalue greater than 1, with other words where the explanatory power[62] or meaning of the combined factor variable is greater than one variable alone. This rule is generally applied in Principal Component analysis and Factor analysis however in the latter case the analyst can chose to calculate the average extracted communalities to determine the eigenvalue cut-off based on which criteria to follow [YONG-PEARCE, 2013, p. 90].
b. Retain factors based on cumulative percentage of variance explained. In Principal Component Analysis the cumulative cut-off is 80% of the total variance. [SZŰCS, 2004, p.441] However, in case of factor analysis this cut-off is determined by the scientific field. In natural sciences the minimum cumulative variance is 95% at the same time in social sciences even 60% can be accepted in some cases. [SAJTOS-MITEV, 2007] Whichever method is chosen by the analyst the interpretability of the selected factors is key. The analyst can include additional factor(s) or exclude selected one if they cannot be interpreted from the research perspective.

3. **Rotation of the factors and quantifying factor loadings**: Following the analysis of the initial extraction of factors the next step is the rotation of factors. The rotation is a linear transformation of the solution to make the interpretation easier. [HATCHER-O'ROURKE, 1994, p.64] The interpretation of the un-rotated factors can be challenging and even misleading due to variables can be associated to factors they should not belong to. In order to help the practical interpretation, the different dimensions or axes of the factors have to be transformed into an optimal position where they have the maximum explanatory power. Following the rotation communalities and the explained portion of the variance stay unchanged only the factor's proportion of the variance is changing. There are two major types of factor rotation available based on the statistical literature:
    a. Orthogonal (varimax, equimax, quartimax) and
    b. Oblique (non-orthogonal – direct oblimin, promax) rotation methods.

    In cost modelling we try to simplify our variable set in order to help interpretation and to reduce the risk of multicollinearity. Therefore, the analyst is advised to use the orthogonal method as the different dimensions or axis are perpendicular to each other in this method which means the factors are uncorrelated. As opposed to the oblique method where the factors are correlated and at the same time the proportion of the variance explained by the factors is at the maximum. [SAJTOS-MITEV, 2007]

    After the rotation the analyst has to analyse how the variables are loading on each factors. In general, the greater the factor loading is the more of the variable's variance is explained by the factor. The determination of whether a variable's factor load is high enough to consider that the variable is actually loading on the given factor can be challenging as the cut-off value itself is determined by a combination of factors such as the number of initial variables, sample size, chosen factor method, etc. Some authors [SAJTOS-MITEV, 2007] put the cut-off value to |0.30| at the same time others [HATCHER-O'ROURKE, 1994] set it at |0.40|. Author of

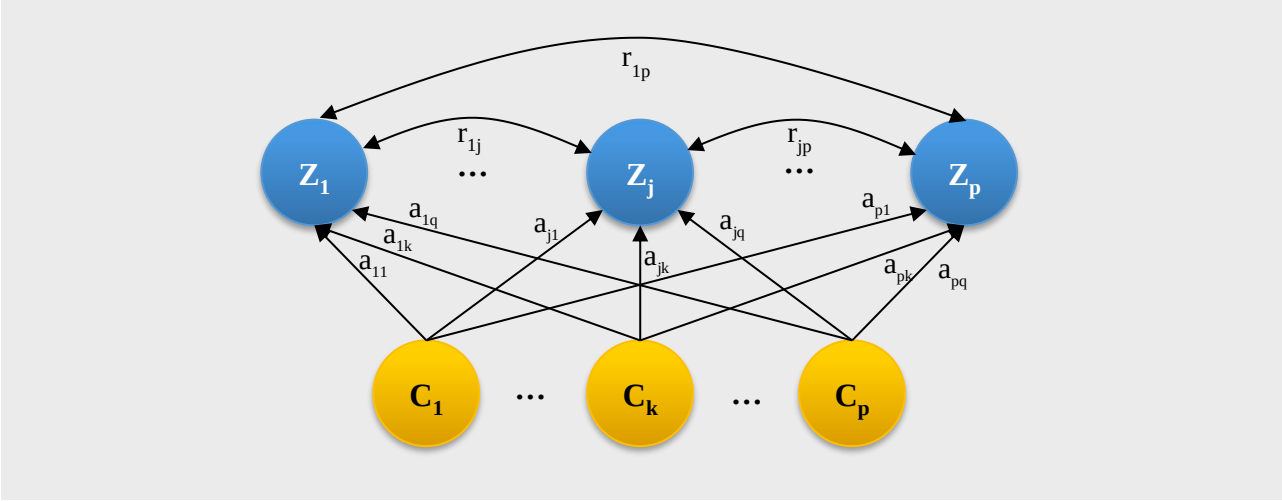
[62] Each observed variable contributes one unit of variance to the total variance. If the eigenvalue is greater than 1 then each unobserved variable explains at least as much variance as 1 observed variable.

current research advises analysts to set the cut-off value to |0.50| with the constant monitoring of the meaning of the factor the variables may load on.[63]

4. **Interpretation of the factors and creating factor scores:** The last step is the definition and naming of the different factors based on the variables loading on them. Although, the factor analysis is following a mathematical/statistical logic the definition of factors requires intuition and a subjective creativity. The analyst should aim to create factor names that cover all composing variables representing their role (weight) in the factor. Once these new unobserved variables are named the analyst has to ensure that the statistical software is recording the factor scores for each observations and the analyst can continue the cost modelling exercise replacing the original correlated variables with the new factors.

The third step is the decision on the application of Principal Component Analysis: Once the analyst has explored the latent variable structure he/she can make a decision whether to use the newly created factors and the variables not linked to any of the unobserved variables or to conduct a principal component analysis where the meanings of all selected variables are condensed to principal components. It is important to highlight that this method can also be used during the regression model building (Stage V.) if the analyst identifies the presence of multicollinearity or would like to optimise[64] the number of variables without major loss of the explanatory power for example.

The fourth (optional) step is the principal component analysis: The PCA method is very similar to the above detailed EFA method. Figure 16 demonstrates the graphical model of the method. As discussed earlier the key difference is the way that PCA accounts for the maximal amount of variance, with other words this method does not differentiate between common and unique parts of the variance.



Source: Compiled by author based on SZŰCS, 2004 p.416

Figure 16: Graphical model of principal component analysis

The linear model of principal component analysis can be written as follows:

$$Z_1 = a_{11}C_1 + a_{12}C_2 + \cdots + a_{1k}C_k + a_{1q}C_q$$

$$Z_j = a_{j1}C_1 + a_{j2}C_2 + \cdots + a_{jk}C_k + a_{jq}C_q$$

[63] Author also suggests to consider factor loadings higher than |0.70| as significant.
[64] In this case the optimisation means variable reduction.

$$Z_p = a_{p1}C_1 + a_{p2}C_2 + \cdots + a_{pk}C_k + a_{pq}C_q$$

where, Z_j is the j^{th} standardised observed variable (X_j); a_{jk} is the coefficient that signifies the importance of the k^{th} factor variable and its weight affecting the change in the value of the j^{th} standardised observational variable; C represents the principal components; r_{jp} is the correlation coefficient to indicate the strength of the relationship between the j^{th} and p^{th} standardised observational variables.[65]

The following tests are required by the analyst to complete the principal component analysis; as the steps are similar to the previously detailed factor analysis author will only list the steps and highlight the differences between the two procedures.

1. **Assessment of the prerequisites**: Similarly to the factor analysis, the analyst has to assess the theoretical basis of the principal component analysis and search for statistical evidence to understand whether the model meets all criteria of PCA.

2. **Initial component extracts and the determination of the number of factors to retain**: Although this step is similar to EFA, the key difference is that in PCA the initial number of components equal to the number of variables from the model. The analyst has similar options when considering which components to retain:
    a. Kaiser Criterion [KAISER, 1960]: eigenvalue > 1 (alternatively analyst can use the Scree test, graphical assessment of the eigenvalues and retain factors before 'elbow');
    b. Cumulative proportion of variance explained (over 80%);
    c. Interpretability: key difference is that not all composite variables (components) have a conceptual meaning due to the underlying assumptions of the PCA (the variance is not split into unique and common variance).

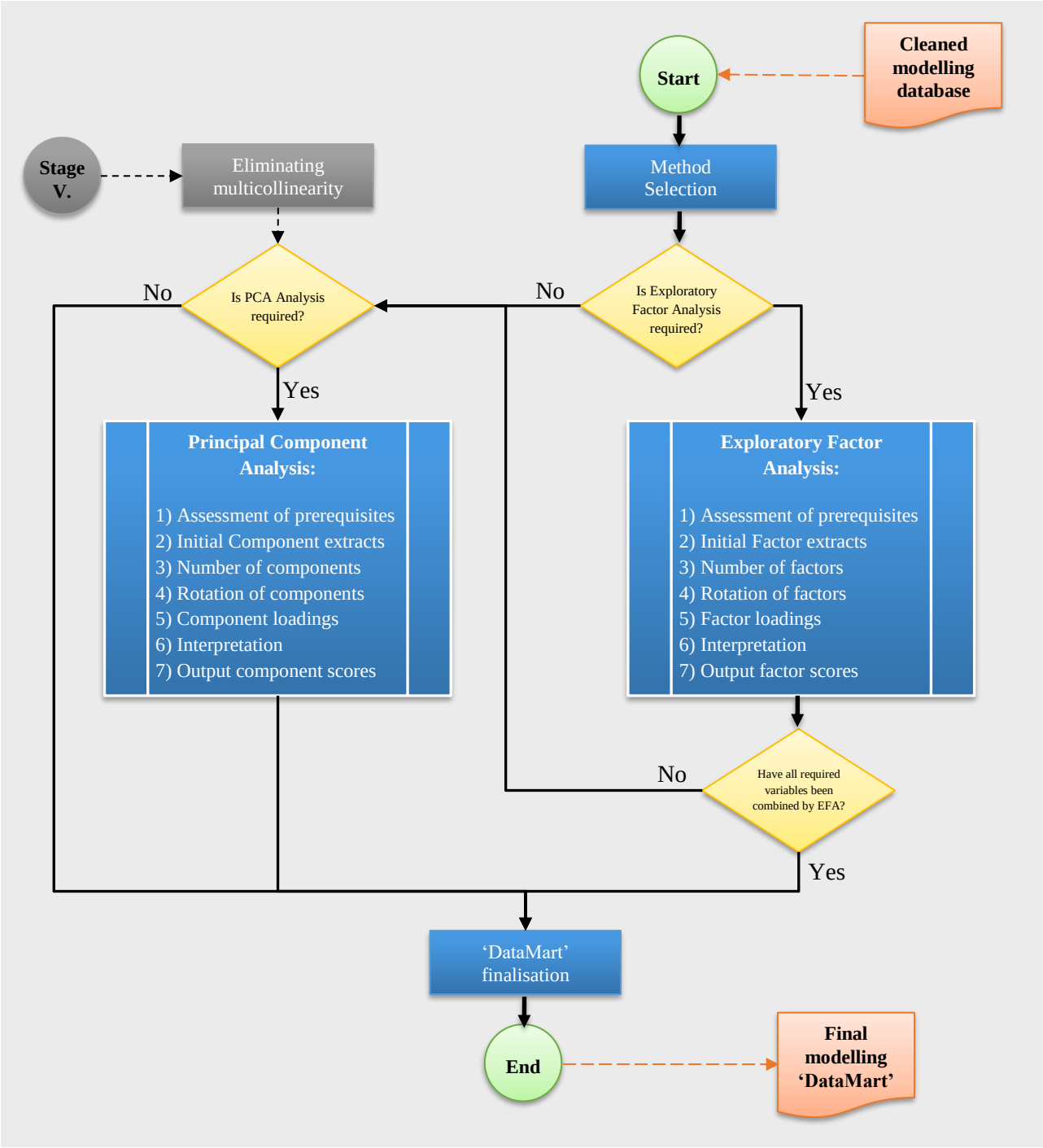
3. **Rotation of the components and the component loadings:** The analyst has the same options (orthogonal and oblique transformation) as at the factor analysis. In the context of PCA the loadings are equivalent with the correlations between observer variables and components. [SUHR, 2005 p.4]

4. **Interpretation of the components and creating component scores:** The last step is the same as in the EFA process. However, the analyst may face situations where there are no straightforward interpretations for all components (especially where the eigenvalue is low). The new component scores (unobserved variables) are ready to be used for linear regression modelling.

The last step is the finalisation of the modelling 'DataMart'[66]: After completing the first four stages of the cost modelling, the analyst can combine the target variable with its cleaned, filtered rationalised (in numbers) predictor variables. In addition to these variables, given the time series nature of the data[67], the analyst is advised to add the time units to the database as a separate variable

[65] For more details on the mathematical algorithm behind the principal component analysis please read SZŰCS, I. (2004): Alkalmazott Statisztika pp. 406-446.
[66] By DataMart the author refers to the database specifically adjusted for the modelling purpose.
[67] For details please see the Managing the specifications of time-series regression section of Stage V. – Multivariate Linear regression modelling

then order the data (ascending order) by the time series. Once the ordering is complete the time units can be replaced with an incrementing number marking the different observations at different time periods. These stages and steps complemented with the conceptual considerations should provide sufficient preparation for a successful cost modelling. Figure 17 summarises the process of reduction of dimensions.



Source: Compiled by author

Figure 17: Process diagram of Stage IV. – Reduction of dimensions

3.3.5 Stage V. – Multivariate Linear regression modelling

As the final analytical 'DataMart' is ready with the identified target and explanatory variables for the selected time-period and frequency, the analyst can ask the question what the difference is between a

simple time series analysis and a regression model based on the same time series data (time-series regression). Regression models require a different mindset from the analyst compared to time series analysis. Time series analysis can identify the trend, the function that describes the change in the value of the target variable over time in a two-dimension space where one of the dimensions is time itself. Even the most complex time series models "can only" follow the target variable over time and estimate a likely value, assuming that the same trend continues. In contrast, regression models aim to explore the relationship through the direct and indirect effects of the explanatory variables to the target variable, with other words the effect of the modelled environment or factors. In this context time is only one of many factors. There are different regression models available[68] depending on the aim of the analysis but in general, in predictive cost modelling, analysts use multivariate time-series regression.

Selection of the modelling strategy: The multivariate regression analysis helps the analyst to understand, describe (and ultimately apply) the complex network of effects between the target and explanatory variables. The regression analysis can be based on one, two, three, four, etc. explanatory variables and the relationship between the target and explanatory variables can be linear and non-linear. It is important for the analyst to understand the implications of using multiple variables before setting the model up. [HAJDU et al., 1998] For example,

a. If the model is intended for an accurate cost prediction or evaluation, then the analyst should include into the model all (meaningful) identified variables either directly or indirectly using aggregated variables (e.g. factor scores from factor or principal component analysis).
b. If the aim of the model is to provide insight to senior management or to analyse specific questions then the analyst should aim to reduce the number of explanatory variables to a level where the interpretation is still feasible without significant loss of explanatory power.

Setting up the linear model: Although the analyst will use a statistical software to complete the model, it is very important to understand the mathematical function and logic of linear regression. This will help the analyst to read and understand the software output of the statistical method and make a decision whether any corrective actions are required or not.

The standard formula of the linear regression is the following:

$$Y=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_ix_i+\varepsilon$$

where, Y is the measured or dependent variable[69];
$x_1, x_2, \ldots, x_i$ are the explanatory, independent or often called as predictor variables;
β_0 is the constant element or the Y intercept;
$\beta_1, \beta_2, \ldots, \beta_i$ are the regression coefficients of the respective predictor variables;
ε is the error or noise, it captures all other factors, which influences the dependant variable other than the x_i.

[68] For example, cross-sectional regression and time-series regression. Cross-sectional regression is when the regression is based on observations from a given time period. Time-series regression is when the regression is based on data linked to a time series. In academia (social sciences, biology, chemistry, etc.) the analyst normally collects the data through experiments, observations, surveys, etc. In a corporate environment the data is normally available in internal or external systems.
[69] In current study it is referred to as target variable.

The main mathematical aim is to minimise the ε error element of the equation. It is achieved when the total distance between the estimated values and the actual values is minimal, which in mathematics is accomplished when the residuals' sum of squares is at the minimum.

$$\sum_{i=1}^{n} e^2 = \sum[y_i - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i)]^2 \rightarrow min$$

As y_i, x_1, x_2, … , x_i are known data elements, the estimation of $\beta_0, \beta_1, \beta_2, \ldots, \beta_i$ parameters are required. The minimum point of the above function is where the partial derivatives of the dependent variable with respect of the various independent variables are 0. After derivation we get n pieces of normal equations where n = i+1 if 'i' represents the number of explanatory variables, with other words the number of normal equations will be equal to the number of unknown parameters. Using one of the following methodologies to calculate / to solve the equations, the value of parameters can be calculated:

- Using simple algebra, we can sort one equation to denote the value of one of the parameters and using this equation we can calculate the value of the other unknown parameters from the system of equations. [SZŰCS, 2004. p.295]
- Using the transformation of the x and y variables;
- Using the rules of linear algebra, [KERÉKGYÁRTÓ et al., 2009] the original logic will be the following:

First we have to create column vectors and matrixes from all known variables.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_k \end{bmatrix} \quad x = \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{i1} \\ 1 & x_{12} & x_{22} & \ldots & x_{i2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1k} & x_{2k} & \ldots & x_{ik} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \ldots \\ \beta_i \end{bmatrix}$$

where, i represents the i^{th} explanatory variable or parameter
k represents the k^{th} observation
ik represents the k^{th} value of the i^{th} explanatory variable
y is a k-by-1 column vector
x is a (i+1)-by-k matrix
β is the (i+1)-by-1column vector of regression coefficients.

Then we have to multiply the transposed x matrix with the original x matrix in one step and in another the transposed x matrix with the y vector.

$$x'x = \begin{bmatrix} 1 & 1 & 1 \\ x_{11} & x_{12} & x_{1k} \\ x_{i1} & x_{i2} & x_{ik} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{i1} \\ 1 & x_{12} & x_{i2} \\ 1 & x_{1k} & x_{ik} \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_i \\ \sum x_1 & \sum x_1^2 & \sum x_i x_1 \\ \sum x_i & \sum x_1 x_i & \sum x_i^2 \end{bmatrix}$$

$$x'y = \begin{bmatrix} 1 & 1 & 1 \\ x_{11} & x_{12} & x_{1k} \\ x_{i1} & x_{i2} & x_{ik} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_k \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_i y \end{bmatrix}$$

where, x' is the transposed x matrix. The original normal equation in linear algebra is the following:

$$x'x\,\beta = x'y$$

Finally, we need to multiply the equation with the inverse matrix of the x'x matrix in order to yield the estimated parameter vector:

$$\beta = (x'x)^{-1}x'y$$

However, the financial analyst will use a statistical software for this calculation and should not need to use these steps directly, it is still important to understand the underlying algorithm for the model evaluation methods (e.g. avoiding multicollinearity is important as the coefficient matrix cannot be inverted if it is singular, etc.) and the interpretation of the output.

Selecting Model Optimisation method: Although making decisions about the inclusion or exclusion of variables purely based on the correlation between the target and predictor variables one by one is reasonably straightforward (as the correlation coefficient is easy to calculate), it may not result an accurate estimation. The main reason for this is the so called multicollinearity. Multicollinearity is the linear relationship between the predictor variables. If multicollinearity is strong[70] then it can cause interference and it is reducing the estimation power of the model. In real life most of the variables have some level of interrelation therefore it is important to;

a) Identify strong, functional relationships and manage the variables before they step into the model (PCA, factor analysis);
b) Use a variable selection method, which takes into consideration the interrelations amongst the predictor variables and includes only the variables into the model that are required for an adequate estimation.

Such models are easier to interpret and the process of variable selection helps the analyst to get an insight into the complex network of variables. Even though, the analyst can choose from multiple methods, current study is focusing on the Forward method, Backward elimination process and the Stepwise inclusion process. [SZÉKELYI-BARNA, 2002]

- The Forward method: introduces the variables to the model one by one. The first explanatory variable to enter the model is where the linear correlation (Pearson correlation) is the strongest with the dependant variable. This is followed by the variable with the strongest partial correlation.[71] The model building is complete when any additional variable would result only a small (insignificant) increment in the explained part of the equation's variance at the significance level of 0.05. Using the forward method, the analyst can understand the sequence of variables which gives an insight to the order of predictors. This can be later used for analysis and strategic decisions.

- The Backward elimination: this method starts with all available explanatory variables included into the model then the variable with the weakest, non-significant partial explanatory

[70] Please see in the 'Testing the model' part the classifications of the multicollinearity.
[71] The inclusion is based on the F statistics calculated for every remaining variable one by one using the following: $F = [\sigma^2(Y) - \sigma^2 (residuals)_{k+1}]/ [\sigma^2(Y) - \sigma^2 (residuals)_k]$, where k is the number of variables already in the model. The variable with the highest F value or with the lowest F significance level will be added. The significance limit is 0.05.

power[72] will be excluded and a new model will be calculated. The iteration continues until only variables with significant explanatory power will stay in the model.

- The Stepwise method: similarly to the forward method, starts with one explanatory variable where the linear correlation is the strongest with the dependant variable. Then introduces a new variable in the same way as stated at the Forward method with one key difference: none of the variables already in the model will definitely stay until the end of the exercise. Before introducing the new variable, the explanatory powers of the variables already in the model are assessed and if the explanatory power of any of the existing model variables has decreased to a level where the t value is not significant anymore ($t<0.05$) then the variable is removed from the model. This method helps to keep only the variables with a significant power and to understand the interrelations of the variables.

It is advisable to complete the model with the chosen method on different significance levels. This will help the analyst to ensure that the correct variables have been selected. Once the initial model building is complete, the analyst has to test whether the model meets all required criteria or any corrections are required before starting the interpretation of the model.

Testing the model: Testing the model is very important as the test ensures that all statistical and modelling criteria have been met and the conclusions made based on the model or the predictions calculated are not under or overestimated. Since the criteria relate to the prediction errors, the tests are assessing the estimated errors and the residuals. The multivariate linear regression modelling is based on four key assumptions:

- **Equal variances:** the errors, ε_i, at each set of values of the predictors, (x_{1i}, x_{2i}, …), have equal variance (σ^2); $Var(Y / X = X_i) = Var(Y / X = X_j) = \sigma^2$
- **Linearity:** The mean of the response, $E(Y_i)$, at each set of values of the predictors, (x_{1i}, x_{2i}, …), is a linear function of the predictors; $E(Y_i) = \beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_iX_i$ $E(\varepsilon)=0$
- **Independence:** The errors, ε_i, are independent
- **Normal distribution:** The errors, ε_i, at each set of values of the predictors, (x_{1i}, x_{2i}, …), are normally distributed; $N(\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_iX_{i,}\sigma^2)$

In order to assess these assumptions, the analyst has to perform a series of tests. Although these statistical tests are designed to support the analyst by providing mathematical evidence, the analyst is also advised to examine the different plots produced by the statistical software packages as sometimes it is easier to understand the interrelation of potential issues rather than focusing on the result of a test. The following tests can be conducted to ensure that the model is adequate to be interpreted:

- **Tests for constant error variance:** The first set of tests the analyst can take are testing whether the errors at each set of values of the predictors have a constant variance. These tests aim to provide evidence for the existence or the rejection of the null hypothesis. The null hypothesis (H_0) says the variances are equal for the different groups of the error ($\sigma^2 = \sigma^2$). The (H_A) alternative hypothesis says the variances are not equal ($\sigma^2 \neq \sigma^2$).

[72] The exclusion is based on the F statistics. $F = [\sigma^2(Y) – \sigma^2 (residuals)_k]/ [\sigma^2(Y) – \sigma^2 (residuals)_{k-1}]$, where k is the number of variables still in the model. The variable with the smallest F value or with the highest F significance level will be removed. The significance limit is 0.05.

- F-Test: This test requires the analyst to split the residuals into two groups. The first group contains the residuals associated with the lowest predictor values and the second group is composed by the errors associated with the highest predictor values. The pooled variances of group one (s_1^2) and group two (s_2^2) are tested whether we have evidence to prove that they not significantly different using the F-Test ($F^* = s_1^2/s_2^2$). This test statistic is distributed according to a F_{n_1-1,n_2-1} distribution, therefore if $F^* \geq F_{n_1-1,n_2-1;1-\alpha}$, then the null hypothesis (H_0) is rejected as the test proves that there is statistically evidence for unequal variances. [SNEDECOR-COCHRAN, 1989 p.223]

- Bartlett's Test: Similarly to the F-Test the analyst (in practice this test is completed by the statistical software) has to divide the residuals into *k* number of groups based on the values of the predictor[73] where group 1 contains the residuals associated with the n_1 lowest values of the predictor until group k where the residuals are associated with the n_k lowest values of the remaining values of the explanatory variable. Then the following equation is applied:

$$B = \frac{(n-k)lnS_p^2 - \sum_{i=1}^{k}(n_i-1)lnS_i^2}{1+\left[\frac{1}{3(k-1)}\left(\left(\sum_{i=1}^{k}\frac{1}{n_i-1}\right)-\frac{1}{n-k}\right)\right]}$$

  The test statistic is distributed based on χ_{k-1}^2 therefore if $B \geq \chi_{k-1;1-\alpha}^2$, then the null hypothesis (H_0) is rejected. [BARTLETT, 1937 p.275] It is important to highlight that the Bartlett's test is highly sensitive to non-normality therefore the Levene Test should be used if the normality assumption is violated.

- The analyst can apply different tests, such as the Modified Levene Test or the Cook-Weisberg Test, etc. [CONOVER et al., 1981 p.352]

- **Tests for Linearity:** Testing linearity is generally easier through the analysis of the scatterplot of residuals versus the estimates or the predictors. However, using the ANOVA table the analyst can test the hypothesis of whether the regression slope is significantly different to 0 or not.
  - Analysis of variance: The linearity assumption is tested by the ANOVA (analysis of variance) procedure:

$$SS_{reg} = \frac{[\sum_{i=1}^{N}x_iY_i - (\sum_{i=1}^{N}x_i)(\sum_{i=1}^{N}Y_i)/N]^2}{\sum_{i=1}^{N}x_i^2 - (\sum_{i=1}^{N}x_i)^2/N}$$

$$SS_{rep} = \sum_{i=1}^{N}Y_i^2 - \sum_{j-1}^{k}(T_j^2/n_j)$$

[73] It is recommended to have at least 25 observations per group.

$$SS_{total} = \sum_{i=1}^{N} Y_i^2 - \frac{\sum_{j-1}^{N}(Y_i)^2}{N}$$

where SS_{reg} is the sum of squares of the regression, SS_{res} is the sum of squares of the residuls due to the variation within repeats of the outcome observations, SS_{total} is the total sum of squares, Y is the dependant variable, x is the predictor, N is the total number of Y observants and n_j is the number of Y repeats for the j^{th} x observation. The test is assessing whether the sum of squares is significantly different from 0. [ARMITAGE et al., 2001 p.314]

- Graphical: The linearity can be also tested by simply using a scatterplot where the chart is showing the residuals (ε_i) on the vertical axis and the estimates ($\hat{y}_i$) on the horizontal axis. Then the analyst can visually assess whether the average of residuals remains close to 0 from left to right. The same exercise can be repeated replacing the estimates on the vertical axis with the predictors one by one.

- **Tests for Independence:** The violation of the independence of error terms (residuals) when using linear regression is a common occurrence especially when the data points were observed in a meaningful sequence such as time series (weekly, monthly, quarterly, etc.). This often means that observations (y_t) at a given point are related to the immediate predecessor (y_{i-1}) therefore the error terms are not independent (ε_t related to ε_{t-1}). In order to determine whether a relationship between the errors exists, the analyst has to assess whether the null hypothesis (H_0) is true and the error terms do not have a positive autocorrelation or the (H_A) alternative hypothesis is applicable, which means the error terms have positive autocorrelation. The residuals are split into independent (μ_t) and lagged residual element ($p\varepsilon_{t-1}$) where the parameter p represents the first order autocorrelation. The tests are determining if the p parameter's value is significantly different to 0 ($-1 \geq p \geq 1$), which means H_0: $p=0$ and H_A: $p \neq 0$

  - Durbin-Watson test: To determine whether to accept or reject the null hypothesis outlined above the Durbin-Watson test is widely used in linear regression analysis. The test is typically test H_0: $p=0$ and H_A: $p>0$ as the error terms tends to have a positive correlation in business and economic applications [NETER et al., 1996 p.497]. The D statistic uses the following equation [DRAPER-SMITH, 1998 p.69]:

$$D = \frac{\sum(\varepsilon_t - \varepsilon_{t-1})^2}{\sum \varepsilon_t^2}$$

The value of D is between zero and four where values close to 2 mean independent error terms. The autocorrelation has two types based on the pattern they follow. Values "close" to 0 suggest the existence of positive autocorrelation which means the errors are clustered around the normal distribution line (e.g. +,+,+,-,-,-,+,+,+). However, values "close" to 4 suggest negative autocorrelation where the error terms alternate (e.g. +,-,+,-,+). The test uses critical value ranges, which are normally based on the number of variables included the sample size and the confidence interval. When the D value is less than the lower limit of the critical range ($D < d_L$) then the null hypothesis

is rejected as there is evidence for positive autocorrelation. In case when D is greater than the upper limit ($D > d_U$) we fail to reject the null hypothesis, with other words the error terms do not have positive autocorrelation.[74] Finally, when the D value is within the range ($d_U > D > d_L$) the test is inconclusive. Once the (positive) autocorrelation is identified it is important to manage it for accurate forecasting.

- Other tests, such as Breusch–Godfrey test or Ljung–Box test, are also available to test the independence assumption and especially whether serial-correlation is present in the model or not.

- **Tests for Normality:** The analyst can choose between different types of tests to determine whether the errors at each set of values of the predictors are normally distributed. Similarly to the other tests, we perform a hypothesis test where the null hypothesis (H_0) is that the errors follow a normal distribution and the alternative hypothesis (H_A) is that the errors do not follow a normal distribution. We are assessing the p-value to determine whether we have to reject the null hypothesis or not.

  - Shapiro-Wilk Test uses the following equation [SHAPIRO-WILK, 1965, p.592]:

$$W = \frac{(\sum_{i=1}^{n} a_i y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad a_i = (a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

    where, y_i is the i^{th} order statistic, $\bar{y}$ is the sample mean, m are the expected values of the order statistics of the independent and identically distributed random variables sampled from the standard normal distribution and V is the covariance matrix of those order statistics.

  - Kolmogorov-Smirnov Test [CONOVER, 1999 p.428] test uses the empirical cumulative distribution and the distribution expected if the data were normal.

$$D = \max(D^+, D^-), where \quad D^+ = \max_i \left({}^{i}/_{n} - F(e_i) \right) \quad D^- = \max_i \left(F(e_i) - {}^{(i-1)}/_{n} \right)$$

    e_i is the i^{th} largest value of the error terms. The statistic is compared to the critical values from a normal distribution and the decision is made based on the significance level (p-value).

  - Although current study is focusing on the above tests, other normality tests are also available: Anderson-Darling Test and Ryan-Joiner Test [NORNADIAH–YAP, 2011].

Managing the specifications of time-series regression: So far we have reviewed the basic model of multivariate linear regression but in economic analysis, especially with cost modelling, we use time-series as the basis of our regression model. The key difference between cross-sectional regression and time-series regression is that the subsequent items or observations in the cross-

[74] Please note that we cannot conclude that the error terms do not have negative autocorrelation either. For this test the analyst needs to use 4-D instead of D for testing.

sectional regression are independent therefore the errors in the linear regression model are independent too unlike in the time-series regression a dependency usually appears in the subsequent data items therefore the errors are dependent too. The phenomena when the errors are correlated over time is called **autocorrelation** or **serial correlations**. As one of the four assumptions of the linear regression modelling is the independence of the errors, the analyst has to understand the potential impact and the required corrective steps to ensure model validity and estimation accuracy. These implications are [KOVÁCS, 1977] are the following:

- The estimated regression coefficients may not have the minimum variance attribute anymore.
- There is an increased chance for the mean square of errors to underestimate the true variance of the errors.
- There is also an increased chance for the standard error of the regression coefficients to underestimate the true standard deviation of the estimated regression coefficient.
- The interval calculation procedure will require modifications due to the violation of the underlying assumptions.

In time series the dependant (y_i) and predictor variables (x_1, …, x_i) are measured over time (t), normally with evenly spaced measurement intervals (weeks, quarters, years, etc.). Due to the nature of this relationship the change in the value of the dependant variable is not only explained by its predictor variables but also impacted by the predecessor value (y_{t-1}) of the same variable. If the change in the dependant variable's value is primarily driven by the previous or lagged values, then the analyst can use an **autoregressive model** where the time series is regressed on previous values from the same time series (y_t on y_{t-1}). Depending on the lag (the time unit difference or gap between the dependant and previous value used) we can differentiate between first-order autoregression AR(1), second-order autoregression AR(2), etc. models[75].

In economic examples the estimation and analysis of time-series questions is not limited to the previous values of the dependant variable, in fact the analyst needs to consider other predictor variables. It presents a potential issue as (due to all variables, including the dependant and explanatory variables are measured as time series) the error term will also be dependent on previous errors. **Regression with autoregressive errors** is when the error term (ε_t) is now made up of two components: autocorrelatied error, which means that a portion of the error can be explained by an error at the previous time ($p\varepsilon_{t-1}$, where $|p|<1$) and white noise or perturbation (ω). $\varepsilon_t = p\varepsilon_{t-1} + \omega$.

Testing Autocorrelation: The analyst is advised to follow the previously detailed tests for independence. Current dissertation applies the Durbin-Watson Test to identify autocorrelation with the assumption that due to the time series nature of the data we will detect positive autocorrelation. Remedial Procedures: As one of the primary assumptions of the regression is violated with the dependent error terms the beta coefficients of the ordinary least squares estimate procedure is unreliable. In order to be able to use the model for cost predicting or management control analysis the analyst will need to apply a remedial solution.

- Cochrane-Orcutt Procedure: This is an iterative process for AR(1) models with the aim to identify p from the previously detailed $\varepsilon_t = p\varepsilon_{t-1} + \omega$ equation. This helps to calculate r, which

[75] The coefficient correlation that measures the linear relationship of the estimate autoregression model and the original model is called autocorrelation function (ACF). In order to identify the order of autocorrelation we need to filter out the linear influence of the random variables, the partial autocorrelation function (PACF) is used.

is the estimated explanatory power of the t-k (where k is the number of lags, in the Cochrane-Orcutt procedure the lag is 1) value of the same variable.

The first step is the transformation of the original regression model

$$\text{from } y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \ldots + \beta_i x_{t,i} + \varepsilon_t$$

$$\text{to } y_t^* = \hat{\beta}_0^* + \hat{\beta}_1^* x_{t,1}^* + \hat{\beta}_2^* x_{t,2}^* + \ldots + \hat{\beta}_i^* x_{t,i}^* + \varepsilon_t^*$$

where, $y_t^* = y_t - r y_{t-1}$, $x_t^* = x_{t,i} - r x_{t-1,i}$, r is the impact estimate of the lagged value and $\hat{\beta}_0^*, \ldots, \hat{\beta}_i^*$ are the estimates regressed on y_t^*. Once the second regression ("to") is complete the Durbin-Watson test will highlight if the autocorrelation still exists in the model and therefore another iteration is required. In order to calculate the r the analyst will need the original and the transformed error terms. The coefficient of the simple linear regression model with the dependant original error terms and the predictor lagged residuals will provide the autocorrelation estimate. Once there is no need for more iterations the analyst needs to transform the regression coefficients back using the autocorrelation estimate: $\hat{\beta}_0 = \hat{\beta}_0^*/(1 - r)$ and $\hat{\beta}_i = \hat{\beta}_i^*$. [COCHRANE-ORCUTT, 1949] In some cases (e.g. extreme[76] positive autocorrelation) the Cochrane-Orcutt method can underestimate p, which ultimately reduces the effectiveness of the prediction. Another, more direct, approach the Hildreth-Lu Procedure follows the same logic but instead of iterations it requires the analyst to test a series of candidate values for p based on the graphical analysis of the pattern of the error terms. [PARDOE et al., 2016]

- First Differences Procedure: in economic examples[77] the value of p is large for AR(1) errors and analyst often make a decision to set p=1 in the transformed model. This procedure is similar to the above detailed Cochrane-Orcutt approach and starts with the transformation, $y_t^* = y_t - y_{t-1}$, $x_t^* = x_{t,i} - x_{t-1,i}$ but in this method the value of r is 1 and therefore it is not part of the transformation anymore. The coefficient transformation is even more simple: $\hat{\beta}_i = \hat{\beta}_i^*$ and $\hat{\beta}_0^* = \bar{y} - (\hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_i \bar{x}_i)$ It is very important in all methodologies to obtain the error terms for every estimated values as these will be required for forecasting. [WOOLDRIDGE, 2001 p.279]

- Advanced methods: The analyst can choose from several advanced methods based on the aim of the modelling exercise. One of the more recent methods is the **autoregressive integrated moving average (ARIMA) model**. This type is often referred to as Box-Jenkins models. These models approach the regression question through the iteration of the following three steps:

    - Model identification and selection: this steps consist of the selection of ARIMA model class (using autocorrelations, partial autocorrelations, etc.);
    - Model estimation: using the maximum likelihood optimisation for the model estimates and the moving parameter of moving average and autoregression;

[76] For example when the Durbin-Watson D test has a value between 0 and 0.5.
[77] Especially when time series data is used but sometimes even in panel data.

- Model checking: the final testing of the fitted model is completed by the analysis of the residuals. The iteration ends when there is no or only minimal improvement achieved with the model fitting process. [BOX et al., 2013 p.93]

Methods, like **exponential smoothing** [SZŰCS, 2004 p.393], **spectral analysis** [WARNER, 1998 p.78] or **generalised least squares** [KMENTA, 1997 p.607] are also available but as current research is focusing on cost and microeconomic strategic questions with the speed and efficiency of the analysis/forecast in mind, author will apply the appropriate methodology for the regression model. It is important to highlight that all of the above described procedures can be performed simply with most of the statistical software packages[78]. Therefore, there is no need for the analyst to follow these steps. However, the understanding of the principles of these procedures will help the analyst to make an informed decision about the model in question.

Managing multicollinearity: Multicollinearity is present when two or more of the explanatory variables in the regression model are moderately or highly correlated. We distinguish two types of multicollinearity: structural (when a variable is a mathematically created from another variable(s), e.g. aggregated variables or proxy variables[79]) and data-based multicollinearity (a common occurrence in observational databases). In Stage III. we already discussed the common variable relationships to avoid in regression analysis and also covered the initial use of the correlation matrix. Although in Stage IV. we have learnt how to implement factor analysis to explore the interrelation of the predictors (and how to use PCA in case we would need the combined explanatory power of all variables), the analyst needs to be able to identify the presence of multicollinearity in case none of the previously described measures solves the issue prior to modelling. Untreated multicollinearity has several effects on the model:

- When (perfectly) uncorrelated predictors are entered into the model their relationships with the dependent variable are unrelated and will show the same coefficient (or slope) as they show when the predictors are separately entered into the model. It means that we always get the same scientific conclusion. Unlike when highly correlated predictors are entered the additional variable will only account for a smaller explanatory power, as the previously entered variable is already partially explaining the same phenomena due to the correlation between the predictors. It means the order of the model entering has an impact on the relationship between the predictor and the dependant variable from the model's perspective. However, there is only one underlying common factor between the two variables that is not measured in the model:

    - Depending on the order the variables entered we can potentially get to different scientific conclusions, which defeats the objective of the model.
    - Furthermore, hypothesis test of the significance of any predictors ($\beta_i=0$) can yield different results due to the order.

[78] For example: SAS Enterprise guide 5.1; Minitab, SPSS v16, etc.

[79] When proxy variables are created in a way that each quality measure is expressed in a separate dummy variable (for example in case of highest completed education is university: yes/no, … is high school: yes/no) then if the analyst would create a variable for all option, the newly created variables would be in a linear deterministic relationship with each other as in case of all new variables being 0 would determine the value of the last variable and vice versa. The analyst has to exclude at least one variable to cease the linear relationship for example.

- The more predictor variables included to the model, the less estimation accuracy will be reached as the degree of freedom is decreasing.

If multicollinearity is still present in the regression model after the analysis of the correlation matrix and the factor analysis, then analyst can use the variance inflation factors (VIF) measure. The calculation is based on the following equation:

$$VIF_i = \frac{1}{1 - R_i^2}$$

where, R_i^2 the R^2 value obtained by regressing the i[th] predictor variable on the rest of the predictors. It shows how many times greater the variance of the estimated coefficient compared to the variance without the multicollinearity.

- If the value of VIF is less than 1 then there is no evidence for multicollinearity;
- If 1 ≤ VIF < 2 then there is a weak multicollinearity in present;
- If 2 ≤ VIF < 5 then there is a moderate multicollinearity in the model;
- The VIF value greater than or equal to 5 suggested a high correlation between the predictor variables. [DOMÁN, 2005 p.25]

Once the presence of still existing multicollinearity is established the analyst can decide: whether to change one of the variables and collect alternative data; increase sample size; exclude one of the variables (based on economic rationale); use a PCA to create a new combined variable. The latter method allows the analyst to keep all variables however makes the interpretation more difficult.

Evaluation of the model fitness (adjusted coefficient of determination): The most common way to measure how well a simple (one predictor based) model explains the reality, with other words how close our estimated values are to the observed values, is through the coefficient of determination or the r-squared value. This metric is the regression sum of squares (SSR) divided by the total sum of squares (SSTO), which determines the percent of the variation in y when predictor x is taken into consideration..

$$r^2 = \frac{SSR}{SSTO} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad or \qquad r^2 = 1 - \frac{SSE}{SSTO} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i is data point of the dependent variable, $\bar{y}$ is the mean of the dependent variable, $\hat{y}_i$ is the estimated data point and SSE is the error sum of squares. The value of r-squared is between 0 and 1 where r^2 = 1 means that all estimated data points fall perfectly on the regression line and $r^2 = 0$ means that the estimated regression line is perfectly horizontal therefore 0% of the variation in y is accounted for by the variation in predictor x. This methodology has limited applications when using multiple variables as predictors. The value of r-squared never decreases when a new predictor is added furthermore the coefficient of determination will suggest that the best possible model is where the number of predictors is equal to the number of observations. It is not optimal as the chance of multicollinearity is higher when a 'large' number of predictor variables are included and the degree of freedom is reduced, which potentially can jeopardise the predictive capability of the model. This highlights the importance to keep the number of variables low without the significant loss of

explanatory power. In order to achieve this goal, the analyst could use a more meaningful measure, the adjusted coefficient of determination, which takes the number of predictor variables into consideration and promotes more competitive model structure. [DOMÁN, 2005 p.17] The analyst can choose from the following different methodologies:

- Theil adjusted coefficient of determination: The value of the metric can decrease with the increase in number of predictors. In some (extreme cases) the value of this indicator can be negative. The primary application is for comparison between different models. [THEIL, 1961]

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

  where n is the total number of observations (or sample size) and p is the number of explanatory variables in the model.

- Other metrics, such as the Schwarz Bayesian Criterion (or SBC) or the Akaike information criterion, can also be used to primarily compare models. [DOMÁN, 2005 p.18.]

Interpretation of the parameters: Once we are confident with the explanatory power of the regression model we can start to interpret the results. The following equation is the output of the linear regression model, which shows our estimates for the values of the Y function based on the variables we included to the model:

$$\hat{Y}=b_0+b_1x_1+b_2x_2+\ldots+b_ix_i$$

where $\hat{Y}$ is the estimate function;
$x_1, x_2, \ldots, x_i$ are the predictor variables;
b_0 is the constant element or the $\hat{Y}$ intercept;
$b_1, b_2, \ldots, b_i$ are the regression coefficients of the respective predictor variables.

When interpreting the model, we have to interpret the regression coefficients:

- b_0 is the Y-intercept, as this is a constant it can be interpreted as the value that the model would predict for Y when all predictor variables $x_1, x_2, \ldots, x_i$ are equal to 0.
- $b_1, b_2, \ldots, b_i$ regression coefficients show the partial effect of the different explanatory variables separately. Although each coefficient is influenced by the other variables in a regression model[80], one coefficient is always showing the partial contribution of the respective variable to the Y value. When x_i is changed by one unit, and assuming all other variables are unchanged then the value of y will change at the rate of the respective regression coefficient b_i. The regression coefficients represent the rate of change of one variable (y) as a function of changes in the other (x_i). The cumulative effect of coefficients will determine the slope of the estimate function.

[80] It is due to the fact that predictor variables are nearly always associated, two or more variabes may explain the same variation in Y. Therefore, each coefficient does not explain the total effect on Y by its corresponding variable as it would if it was the only variable in the model. Rather, each coefficient represents the additional effect of adding that variable to the model, if the effects of all other variables in the model are already accounted for. Therefore, each coefficient will change when other variables are added to or deleted from the model.

Testing Research Hypotheses: As previously discussed the analyst can build the regression model for different reasons (predictive reasons, control purposes, inferential or theoretical reasons, etc.). If the aim of the model is to test whether one, a subset or all parameters associated to a predictor are 0 or not, with other words whether a variable explains the variance of the dependant variable at all then the model can be used for hypothesis testing [KEHL-SIPOS, 2010]:

- Is the assessed explanatory variable significantly (linearly) related to the dependant variable?

  $H_0 : \beta_1 = 0$ (the null hypothesis means that the variable is not significantly related)
  $H_A : \beta_1 \neq 0$ (the alternative hypothesis means that the variable is significantly related)

- Is the selected subset of explanatory variables significantly (linearly) related to the dependant variable?

  $H_0 : \beta_2 = \beta_3 = 0$
  H_A : at least one $\beta_j \neq 0$ (for j = 2, 3)

- Does the regression model contain at least one explanatory variable significantly (linearly) related to the dependant variable?

  $H_0 : \beta_1 = \beta_2 = \beta_{max} = 0$
  H_A : at least one $\beta_j \neq 0$ (for j = 1, 2, max)

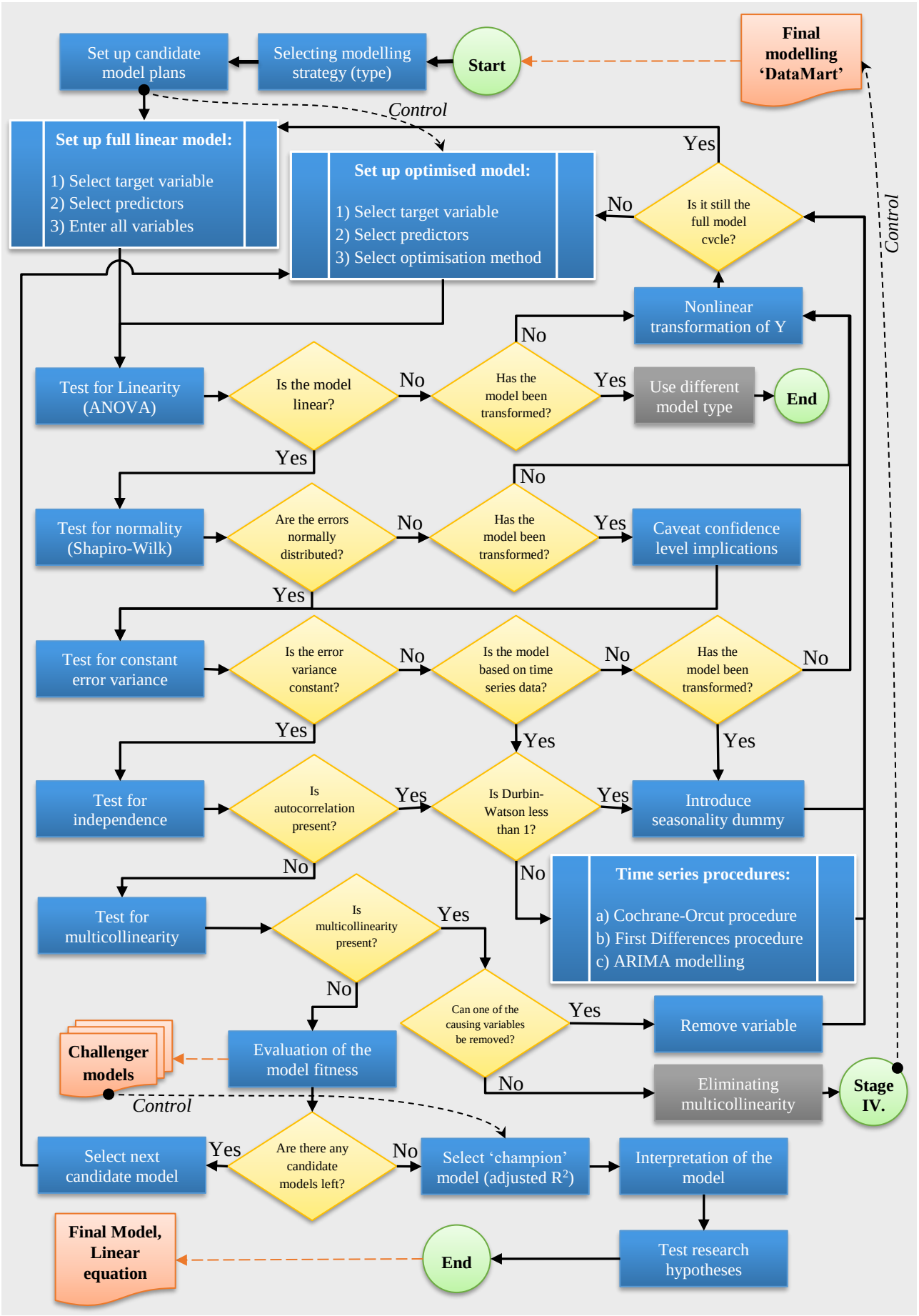
Then the hypotheses are tested with the general linear F-Statistic. Using the following equation and the F-distribution probability table the analyst rejects or keeps the null hypothesis based on the p value. As previously discussed current research applies p = 0.05 as the significance limit.

$$F^* = \left(\frac{SSE(R) - SSE(F)}{df_R - df_F}\right) \div \left(\frac{SSE(F)}{df_F}\right)$$

where SSE(R) is the error sum of squares of the reduced model (model without the assessed explanatory variables) and SSE(F) is the error sum of the full model;
df_R and df_R are the respective numbers of error degrees of freedom associated with the reduced and full models.

With the help of these hypotheses test the analyst can find statistical evidences for the existence of relationship between selected explanatory variables and the dependant variable.

Model building strategy: The steps detailed above are part of a repetitive process where the aim is to find the most accurate and effective model for the selected modelling purpose. Although there are several model building strategies, the analyst can follow the strategy outlined on Figure 18: 1) define strategy (prediction, theoretical, control or interferential analysis); 2) define full model (a model where all predictor variables are included), then define challenger models (models with subsets of variables with an optimisation model strategy) to test the original analytical assumptions; 3) complete full model; 4) complete model and goodness-of-fit evaluation; 5) re-iterate process with challenger models; 6) use the adjusted R^2 to select 'champion' model; 7) interpret the model; 8) prepare model for forecasting and linear programming. Although the major steps of the multivariate regression model building process are controlled by statistical rules, it is important to highlight that the capability of the analyst to put the model into context is vital for cost predictive solutions.



Source: Compiled by author

Figure 18: Process diagram of Stage V. – Multivariate Linear regression modelling

3.3.6 Stage VI. – Cross-Validation of the model

Once the final model is identified, the analyst should ensure that the application of the modelling technique was correct and in fact the model is suitable for its purpose especially if the intention is to use it for cost planning. The analyst has multiple options depending on the availability of control data.

In case **the analyst can obtain new observations** or observations from the past that have not been used for the model building the analyst can choose from the following options:

a) **Refit the model on the new dataset:** In this case the analyst can evaluate the different characteristics of the model (e.g. mean, standard error, estimated regression coefficients, etc.) With the use of a statistical test the analyst can determine whether these characteristics are significantly[81] different or not.
b) **Predict values based on existing dataset and compare the prediction against actuals:** The analysis of the prediction errors (the difference between the predictions and the actual values), using the mean squared prediction error, can help the analyst to understand how well the model predicts the new data points.
c) **Test the mean squared prediction error (MSPE) with challenger models:** Using the challenger models from the model building stage the analyst can test all models against actuals and calculate the MSPE for all challenger and champion model. The test should show that the final model has the smallest MSPE value amongst all assessed models. [ARLOT-CELISSE, 2010, p.69]

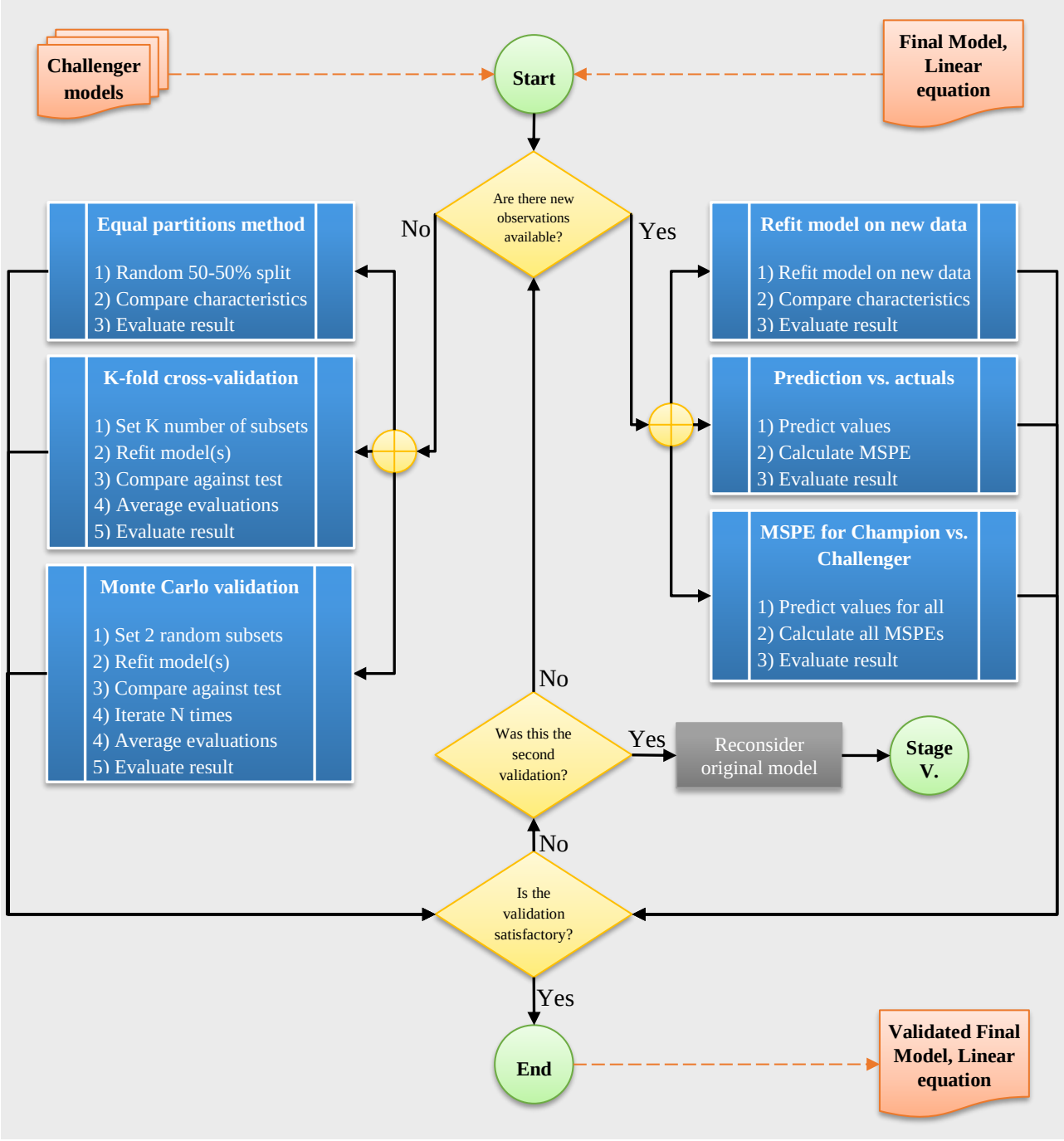
It is more common that there are **no new observations available** at the time of modelling and therefore the analyst should use the final model to set up the modelling and testing partitions:

a) **Refit the model on validation partition:** When new observational data is not available the analyst can partition the existing data into two data partitions: a training (or model building) and a validation partition. The simplest way to do it is to randomly split the data in half[82] however it can present some difficulties when using time series data.[83]
b) **K-fold cross-validation:** This method follows the same model refitting and evaluation logic as detailed above. The key difference is in the partitioning. The process starts with the random distribution of the data into k number of subsample groups of which 1 group is the test group and the rest of the groups (k-1) are validated against this test sample set. The results of the cross validations are averaged in order to produce a single estimation of the model accuracy. [BENGIO-GRANDVALET, 2003]
c) **Monte Carlo cross-validation:** The key difference compared to the k-fold method is that the training and the validation subsets are randomly selected in every iteration. The results of each iterations are averaged for a single evaluation. On the one hand the validation is not dependent

[81] Based on F-statistics for example.
[82] The sufficient minimum is to have 6-10 observations per potential predictor variable in the training set. [PARDOE et al., 2016] If it is not available then the analyst can choose to split the dataset into training versus validation with a different ratio (e.g. 60% vs. 40%) in order to satisfy the minimum criteria
[83] The analyst is advised to split the data into a training and validation set once the lagged variables have been already introduced to the modelling dataset.

on the number of iterations. On the other hand, due to the random selection it is possible that some observations are never tested. [PICARD-COOK, 1984]

In general, one of the above detailed validation methods should provide sufficient evidence about the model validity to continue cost planning. In case the result of the cross-validation is ambiguous the analyst can choose another validation method. In case the second validation would result the same conclusion, the analyst needs to review the original modelling process (Stage V.) and understand if there are specific circumstances (autoregressive errors or outliers in the) that have not been taken into consideration during the model building. If the cross validation is satisfactory then the analyst can proceed to the last stage. Figure 19 demonstrates the process of cross validation.



Source: Compiled by author

Figure 19: Process diagram of Stage VI. – Cross-Validation of the model

3.3.7 Stage VII. – Cost prediction and intervals

The last stage of the predictive cost modelling is the cost prediction and the application of the linear equation. It consists of three main steps depending on the question the analyst needs to answer: the application of linear equation to predict new value; calculation of the prediction interval for predicting new interval; and calculation of confidence interval for estimating the mean.

1. **Predicting new value:** Once the analyst has completed all mandatory steps of the regression model building for cost prediction, the linear equation can be used to predict new values. The analyst should collect the forecast, plan or target values for the variables kept in the final regression model and in any principal component or factor included, which finally made it to the regression equation. The analyst can choose to calculate the future values using any spreadsheet or the statistical software. In case of autoregressive errors, it is very important for the analyst to obtain the original error structure so it can be used to adjust forecasted value based on the seasonal effect. It is less important when the lagged response variable is used as a predictor (e.g. in the first order autoregression model). [PARDOE et al., 2016]

2. **Calculation of the prediction interval:** When the analyst is looking to answer the question around a specific predicted value, it should be noted that the actual value may differ from the prediction due to the imperfections (errors) of the model. Therefore, the analyst is required to provide a range at a given confidence level in which the actual value is expected. For example, when the analyst is trying to predict the success of a specific marketing campaign through the potential number of new customers recruited based on the media spend and other factors then the analyst should include the predicted value and a prediction interval at a given confidence level (e.g. 95%, 99%, etc.) into the answer. In order to calculate the interval, the analyst will need the degrees of freedom and the predicted value plus (or minus) the standard error adjusted by the t-multiplier at a given confidence level.

$$\textit{predicted value} \pm \textit{(t-multiplier} \times \textit{standard error)}$$

$$\hat{y}_h \pm t_{(\frac{\alpha}{2},n-p)} \times \sqrt{MSE + [se(\hat{y}_h)]^2}$$

where $\hat{y}_h$ is the predicted value when the predictor values are $X_h = (1, X_{h,1}, X_{h,2}, \ldots, X_{h,p-1})^T$
$t_{(\frac{\alpha}{2},n-p)}$ is the t-multiplier or t-value for a two-tailed t-distribution and
$\sqrt{MSE + [se(\hat{y}_h)]^2}$ is the standard error of the prediction.

Most of the statistical software offer the calculation of the prediction interval (PI) therefore the analyst should be familiar with the calculation method mainly to support the interpretation.

3. **Calculation of confidence interval for the mean response:** The analyst can also answer other questions not related to a specific event but to the mean of the same type of events based on the regression model. For example, when the analyst is trying to answer the question around the mean success (average number of customers recruited) of marketing campaigns at a given spending level, the result is still subject to the errors of the model. However, the analyst can provide the confidence interval around the mean instead of an interval for a single

point / value. The logic of the calculation is the same as above with the difference in the calculation method of the standard error.

$$\hat{y}_h \pm t_{(\frac{\alpha}{2},n-p)} \times \sqrt{MSE(X_h^T(X^TX)^{-1}X_h)}$$

where $\hat{y}_h$ is the predicted value when the predictor values are $X_h = (1, X_{h,1}, X_{h,2},\ldots, X_{h,p-1})^T$

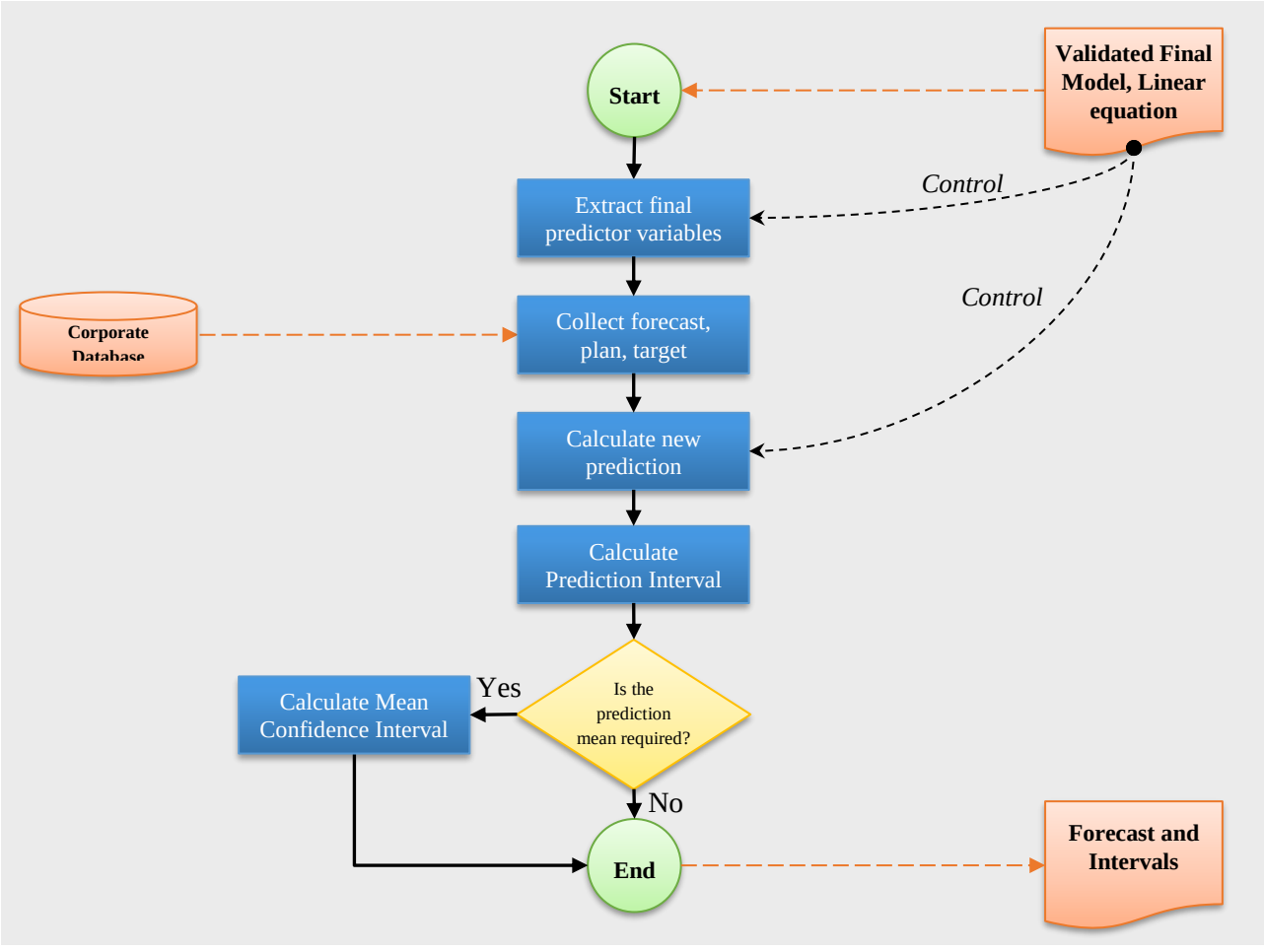
$t_{(\frac{\alpha}{2},n-p)}$ is the t-multiplier or t-value for a two-tailed t-distribution and

$\sqrt{MSE(X_h^T(X^TX)^{-1}X_h)}$ is the standard error of the prediction

Although the two formulas look almost identical, the key difference is the additional mean square error (MSE) in the equation of the prediction interval. This is due to the fact when calculating the confidence for the estimated mean, the mean at a given point is known as opposed to the unknown mean in the calculation of the prediction interval. The unknown mean should be estimated with the variance but as the variance is not known, it can be estimated with the MSE, which yields the addition mean square error in the formula resulting that the prediction interval is generally wider than the confidence interval for the estimated mean.

It is important to note that prediction and confidence intervals are only effective if the regression model has met all four important criteria (linearity, independence, normal distribution, equal variances). [GRANGER, 1986]

The following figure summarises the last stage of cost modelling.



Source: Compiled by author

Figure 20: Process diagram of Stage VII. – Cost prediction and intervals

"Everything should be made as simple as possible, but not simpler."

Albert Einstein (1879 – 1955)

4. RESULTS

Using the data collected from various sources with the application of the methodology outlined in Chapter 3 author is presenting the detailed calculations and research result in the current Chapter with the intention to test all five main hypothesis groups. This section has four main parts: 1) The foundations of the controlling methodology improvement; 2) Improved Planning and Evaluation: Predictive Cost Modelling; 3) Lessons learnt from new method: key factors that determine bank marketing success; 4) The potential: evaluation of the new model. The Chapter is designed around the research aims with keeping not only the scientific community in mind but practical finance and planning professionals too.

4.1 THE FOUNDATIONS OF CONTROLLING METHODOLOGY IMPROVEMENT

In Chapter 4.1 author is using the literature review (from Chapter 2) to understand if existing cost planning methods are dominantly non-statistical based procedures and therefore to prove that there is room for another analytical/statistical based cost forecasting method. Using the collected data from a key publication and a research database author is testing the first two hypotheses (H1, H2) and aims to establish the maturity of the combined literature of management control and accountancy harnessing big data technology and data science.

4.1.1 Potential for improvement: existing management control cost planning processes

Author is using a systematic approach to test the first hypothesis (H1), which states that *existing cost planning, especially marketing campaign cost planning, methods are dominantly non-statistics based approaches.* During the literature review, author has evaluated and categorised the cost planning methods (or marketing cost planning methods where specific literature was available). Current dissertation recognises a cost planning method statistical/mathematical if it involves the application of probability calculations, multivariate methods (principal component or factor analysis, cluster or discriminant analysis) or regression analysis. The toolset of descriptive statistics and indexes are classed as the non-statistical (heuristic) methods for the purpose of current research. The following four categories have been artificially created in the dissertation:

1. Non-statistical methods only: 100% of the described methods in the selected scientific work of the author are heuristic with no mention of statistical methods.
2. Dominantly non-statistical methods: 50-99% of the described methods in the selected scientific work of the author are heuristic with limited (1-49%) mention of statistical methods.
3. Dominantly statistical-methods: 50-99% of the described methods in the selected scientific work of the author are statistical with limited (1-49%) mention of heuristic methods.
4. Statistical methods only: 100% of the described methods in the selected scientific work of the author are statistical with no mention of heuristic methods.

Author considers H1 Hypothesis proven if over 75% of the selected literature is categorised as non-statistical or dominantly non-statistical method. Table 13 summarises the results of the categorisation of the processed literature.

Table 13: Categorisation of cost planning methods in the literature

	Mentioned planning methods	Direction	Category
ZÉMAN, 1998	Classical, Process based, Activity based and Target costing	Hungarian	Category 1
CLARK, 1999	Classical, metrics, indexes	US/UK	Category 1
MABBERLEY, 1999	Activity based costing	US/UK	Category 1
BODA AND SZLÁVIK, 2001	Classical and Activity based costing	Hungarian	Category 1
HORVÁTH, 2004	Activity Based Costing, Budgeting, Classical cost planning	Hungarian	Category 1
JÓZSA et al., 2005	Classical, Indexes, trends	Hungarian	Category 1
FRANCSOVICS, 2005	Classical, Management Planning	Hungarian	Category 1
FARRIS et al., 2006	Classical, forecasting based	US/UK	Category 2
GYULAINÉ-JÁGERNÉ, 2006	Classical and Activity based costing	Hungarian	Category 1
KÖHLER, 2006	Classical, Activity based, Target costing, Probability calculations	German	Category 2
KÖRMENDI AND TÓTH, 2006	Classical, Process based, Activity based and Target costing	Hungarian	Category 1
REINECKE AND FUCHS, 2006	Heuristic and Analytical approaches	German	Category 2
SZÓKA, 2007	Classical, Process based, Activity based and Beyond Budgeting	Hungarian	Category 1
STRAUSS, 2008	Pragmatic and Analytical approaches	US/UK	Category 2
KANDIKÓ AND LEHEL, 2012	Metrics, indexes and probability calculations	Hungarian	Category 2
HAJDÚ, 2013	Classical, Indexes	Hungarian	Category 1
MEFFERT et. al., 2015	Classical and dynamic planning	German	Category 2
RIEG, 2015	Classical, Advanced, Beyond Budgeting	German	Category 2

Source: Compiled by author

11 (61%) out of the 18 authors have only described non-statistical planning method in their publications. Mainly Classical, Activity- and Processed based methods have been detailed. The rest 7 authors have been classed as dominantly non-statistical based method provider. Beyond Budgeting and Probability Calculations have been common statistical methods however the methods described were still dominantly classical or activity based. Although this finding provides evidence to the validity of H1 hypothesis due to the sample size and the selection method (total population unknown, not random sampling etc.) author, acknowledges the result as partial confirmation of the hypothesis.

4.1.2 Enabler of improvement: The maturity of the literature of data science and big data

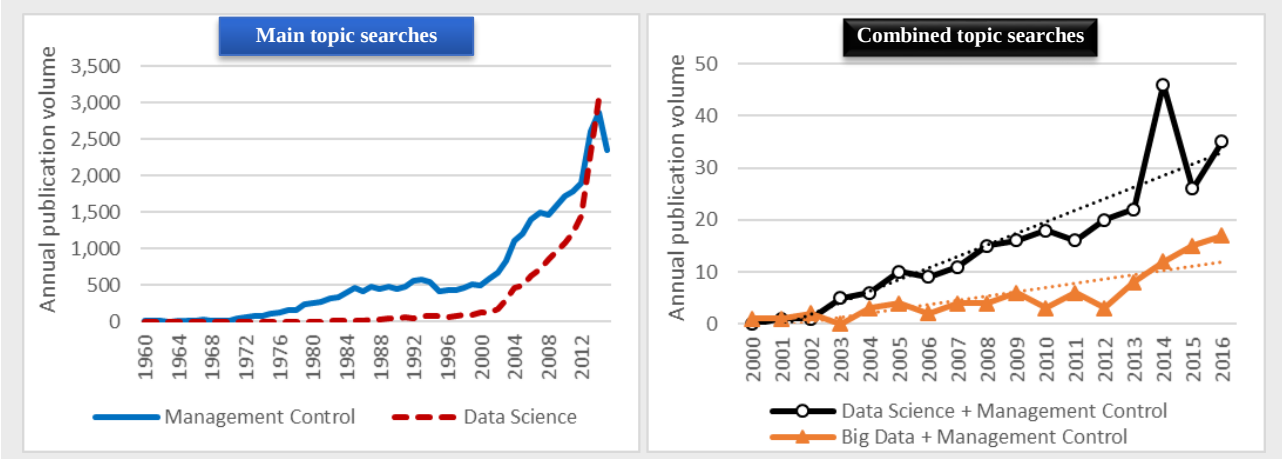
As we have seen from the literature review, the Big Data driven data science is a new branch of science. The first mention of Big Data technology (distributed storage and processing of large datasets using computer clusters) was in 2003 when Google released a whitepaper on a new File System. [GHEMAWAT et al., 2003] Although statistical methods applied by data science have long existed before big data, the technical requirements and the possible practical application of machine learning for example, developed in parallel with the technology. It raises a question whether the literature of the field, especially the combined literature of management control and data science, is still considered new or already matured (well-researched and well-published) and therefore new researches, such as current research, would only have marginal contribution.

Author's hypothesis is that *the field of management control systems, harnessing the Big Data technology and data science is new and has only been emerging from the literature in the last decade.* The literature of management control harnessing Big Data technology and data science is still developing (it shows an upwards trend with no evidence of a plateau yet).

Author for the hypothesis testing is using the ProQuest[84] comprehensive search engine of publications (articles, features, news, dissertations, conference papers and proceedings etc.). The search engine is sourcing publications from three main databases: ABI/INFORM Global (Business, Management and Trade - scholarly and trade journal articles, dissertations, SSRN working papers, business cases and global and trade news since 1971); ProQuest Business Collection (Largely full-text database combining scholarly journals, trade publications, dissertations, working papers, market reports, newspapers, and other relevant sources in business and the social sciences since 1951); and The Vogue Archive (The full contents of the US version of the Vogue magazine (US edition) since 1892 with monthly updates for new issues). There are the following four search criteria set[85]:

1. Management Control publications: publications with the exact text of "management control" in the publication title and/or in the key words and/or in the abstract.
2. Data Science publications: publications with the exact text of "data science" in the publication title and/or in the key words and/or in the abstract.
3. Data Science and management control combined publications: publications with the exact text of "management control" AND "data science" in the publication, key words, abstract.
4. Big Data and management control combined publications: publications with the exact text of "management control" AND "big data" in the publication, key words, abstract.

The data shows the monthly publication volumes of the different groups where the first (management control publications) category serves as reference (total publication volume since 1960 is 36,091).



Source: Compiled by author based on annual publication volume counts from the ProQuest publication database

Figure 21: Data Science and management control combined annual publication volumes

Based on the above described methodology author can confirm that the field of management control systems harnessing the Big Data technology and data science is new and has only been emerging from the literature in the last decade as 91% (234 out the total of 257) of the data science and management control combined publications and 93% (80 out of the total of 86) of Big Data and management control combined publications have been published between 2006 and 2016. The fitted linear trend function of the respective total annual publication volumes shows a significant positive trend, which proves that the literature is still developing therefore the current research has higher impact (1/257) compared to publications in the general management control field (1/36,091).

[84] For more information please visit: http://www.proquest.com/
[85] The search is completed only within Books, Conference Papers, Dissertations and Theses, Reports, Scholarly Journals and working papers. No language filter has been applied.

4.2 THE PREDICTIVE COST MODEL

In Chapter 4.2 author applies the methodology outlined in Chapter 3.3 in order to test the hypotheses related to the predictive cost modelling (H3, H4.1, H4.2, H5.1). The data for hypothesis testing is a simulation database generated based on data from an existing retail bank operating in the United Kingdom. The de-sensitised data preserved the original relationship between the assessed variables but not reflecting the actual values of the institution (as detailed in Chapter 3.2).

4.2.1 Stage I. – Identification of the strategic aim

The first step is to identify the strategic aim. As detailed previously the analyst can use several methods for this. Author has aimed to understand the strategic priorities not only for the financial institution, on which the simulation data is based, but also for the other largest banking groups to provide a more representative view of priorities. The data is collected from publicly available annual statements of the financial organisations that held over 60% of the total asset (excluding the Central Bank) in the United Kingdom at the end of 2015. Table 14 summarises the key strategic priorities of the top five[86] financial institution in the United Kingdom for 2016 and 2017.

Table 14: Common strategic priorities in the financial industry

	HSBC Holdings[87]	Barclays PLC[88]	RBS Group[89]	Lloyds Group[90]	Santander UK[91]
Total asset[92]	727.9 bn GBP 12% of market	1120.9 bn GBP 18% of market	815.4 bn GBP 13% of market	806.7 bn GBP 13% of market	281.4 bn GBP 4% of market
1	Reduce risk weighted assets across the group	1st place Relationship Net Promoter Score© (NPS) vs. peer sets	Maintain bank CET1 ratio of 13%.	Creating the best customer experience	Customer Loyalty and market share growth
2	Optimise HSBC's global network	Top 3 ranking of wallet share or customer satisfaction with priority clients	Narrow the gap to No.1 for NPS in every primary UK brand.	Becoming simpler and more efficient	Operation and digital excellence
3	Set up a ring-fenced bank	Sustained engagement of colleagues score	Reduce operating expenses by £800m.	Delivering sustainable growth (market leading in current account and mortgages)	Consistent and growing profitability and a strong balance sheet
4	Deliver 4.5-5.0 billion USD cost savings	Conduct Reputation (6.5/10 YouGov survey)	Net 4% growth in PBB and CPB customer loans.	Building the best team	Live The Santander Way through our behaviours
5	Deliver revenue growth above GDP	Adjusted Return on Equity	Raise employee engagement to within two points of GFS norm.	-	Support communities through skills, knowledge and innovation

Source: Compiled by author

[86] Based on the total consolidated UK asset value at the end of December 2015
[87] http://www.hsbc.com/~/media/hsbc-com/investorrelationsassets/hsbc-results/2015/annual-results/hsbc-bank-plc/hsbc-bank-plc-annual-report-and-accounts-2015.pdf
[88] https://www.home.barclays/annual-report-2015.html
[89] http://investors.rbs.com/~/media/Files/R/RBS-IR/annual-report-2015/strategic-report-2015.pdf
[90] http://www.lloydsbankinggroup.com/globalassets/our-group/responsible-business-2016/download-centre/2015_lbg_annual_report_v3.pdf
[91] http://www.aboutsantander.co.uk/documents/Santander%20UK%20Group%20Holdings%20plc%202015%20Annual%20Report.pdf
[92] In billion GBP. Sourced from the annual statements above. % is based on total asset of 6,289.7 bn GPB at end of December 2015. Source: http://www.bankofengland.co.uk/statistics/Documents/bankstats/2016/oct/tabb2.1.xls

Most of the assessed large banks focus on all elements of the balance scorecard from the customers through colleagues to investors. An increased attention around customer satisfaction and net promoting as well as around efficiency and cost rationalisation can be identified. Furthermore, almost every banking groups set aims around customer growth and market share growth (highlighted with blue colour in Table 14). Although we can identify other overarching common aims (e.g. establishment of ring-fence banking, etc.) from the marketing control perspective, author is focusing on common strategic priorities that are also in the scope of market management, with other words strategic aims that "*target markets and getting, keeping, and growing customers through creating, delivering and communicating superior customer value*" [KOTLER-KELLER 2012, p.5].

Although marketing management covers a wide range of product design and management such as brand building and sales oriented activities (amongst other tasks), most of the assessed institutions focus on customer and market share aims in their 2016 priorities. Some of the identified customer and market growth related priorities are generic and aim either a certain ranking of the wallet share[93] and customer loyalty[94] or market leading positions and market share growth. Other institutions have specific market position related targets such as targeting a net percentage market growth in certain segments. Even though marketing campaigns can aim several things ranging from brand awareness through customer communication to encouragement of product/service trial and customer recruitment, author considers the above detailed examples and logic adequate to choose the customer volume (increment) and market share variables as key strategic aims or strategic variables for the predictive campaign cost modelling. It is important to mention that there are no conflicting principles or methodological reasons why the cost prediction should not aim any of the above mentioned other campaign priorities. As long as the predictor variable consideration and adequate preparation has been completed for the specific objective, the analyst can choose different target variables. Current dissertation is focusing on the customer volume and market share increase targets.

4.2.2 Stage II. – Identification of the contributing variables

The second stage of the predictive campaign cost modelling starts with the **initial identification of (all) predictor variables**. As outlined in the methodology, the initial aim is to identify all potential predictors regardless data availability or the consideration of multicollinearity, degree of freedom, etc. at this stage. The logical consideration of all explanatory variables will help the analyst to prioritise variables before model building, evaluation and model optimisation. The analyst has multiple options to explore all predictor variables (literature review, qualitative research or 'controlled brainstorming' in a corporate environment, application of previous models). Author has chosen a combination of two options; primarily mapping the network of variables based on the interview based qualitative research[95] complemented by literature review. Using the above detailed methods the following four (main) categories of initial predictor variable groups have been created.

[93] The term 'wallet share' refers to the customers with multi banking relationships. A financial organisation can aim to have any relationship with a customer, with other words 'to be part of the customer's wallet', or to turn this relationship to primary where the bank aims to offer services to the customer where the customer considers the relationship with the financial institution primary (as regular income and key payments are managed from one provider's account).
[94] Customer loyalty is considered as a form of customer relationship where the customer either holds multiple active products of one provider or regular income and key payments are managed from one provider's account.
[95] For details please see Chapter 3.2.

Table 15 summarises the initial variable groups and their categorisation. The detailed list of all initial variables (going down to level 4) can be found in the appendices (A5).

Table 15: Main predictor variables groups

Group	Level 1	Level 2	Level 3
Group 1	General Operations (13 variables)	Infrastructure	Staff, Network size
		Operations and Service Quality	Branch Appointments, Telephone Distribution Appointments, Website visits, Customer Satisfaction
Group 2	Product and Market Share (94 variables)	Product Opening by channel	Current Account, Credit Card, General Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans
		Market Share by Product and Channel	Current Account, Credit Card, General Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans
		Market Competitiveness	Savings rate competitiveness, Press, Current Account
Group 3	Marketing Activity (56 variables)	Marketing Activity	Sponsorship Event, Direct Mailing Campaign, TV Campaign
		Media Spend by Product Type	Total, Current Account, Credit Card, General Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans
		Media Spend by Media Type	Cinema, Digital, Direct Mail, Outdoor, Press, Radio, TV
		Brand Power	Brand Awareness, Brand Warmth
Group 4	External Factors (22 variables)	Economic Environment	Unemployment rate, Gold Price, Base rate, Avg. Bond rate, Avg. Mortgage rate, Inflation, Median house price, CCI, BCI, GDP
		Weather	Average Temperature, Average Rainfall
		Seasonality	ISA season, Student season, Bond season
		Events	Working day counter, Sport Event flag, Political Event flag

Source: Compiled by author

In the following section author briefly describes the variables groups from the initial assessment in order to provide the required understanding for the next steps (potential reduction of dimensions and model building):

Group 1 – General Operations: The first group of variables are focusing on the internal process and infrastructure of the organisation. Based on the interviews it became clear to the author that the general operations of the financial institution should have an impact on the target variables. The underlying assumption is that the size of the organisation and the amount of appointments the frontline staff manage as well as the quality of these operations have a direct impact on the volume of new customers recruited or the market share level reached in certain product campaigns. The interviewed industry professionals suspected this relationship to be linear and directly proportional, with other words the larger the organisation is or the 'better' the operation quality is the more customers are recruited. This is not only driven by the fact that a larger organisation can process more applications but also the size and infrastructure of the organisation can reach out for a larger potential customer audience as the larger frontline staff volume can spread their experience, corporate message faster than at a smaller organisation. The general operations group include the following sub-groups:

- ***Infrastructure***, which has two main components: number of frontline staff (branch based and telephony distribution based separately) per chosen period and the number of branches with the percentage of branch share when compared against all competitors or key peers.

- ***Operations and service quality***, which has multiple variables: number of branch appointments booked and completed; number of calls handled (product or service purchase related); number of website visits; number of service quality related complaints per chosen period and customer satisfaction. For the latter metric, the analyst has multiple options. In case the company subscribes for a regular survey service of an external market and user experience research agency such as GfK[96] for example then (with the consent of the agency) the data can be used for the predictive cost modelling exercise. The benefit of the external data is the availability for competitors and the consistency of the methodology across all peers. The downside is the sample size (as not all surveyed customer will necessary have an interaction with the financial institution in question) and the frequency (currently this service is offered on a monthly or quarterly basis, which would not necessarily reflect mid-month campaign activity). The alternative option is when the organisation conducts its own research. The selection can be based on a panel of customers or certain events such as a transaction at counter, which can trigger a survey. It was suggested by the interviewed industry professionals that for the cost modelling the last options is the most effective as that is measuring the performance of the processes and operational activities from the customer's perspective.

- ***Risk appetite***: The last set of variables are grouped around the risk appetite of the financial institution. It is an important measure as it reflects both the acquisition strategy and the application quality for the financial organisation which ultimately impacts the new customer recruitment ability and the market share change potential.

Group 2 – Product and Market Share: The second group of variables are clustered around the elements of the marketing mix, the 'classic four Ps[97]. It is focusing on variables related to the product propositions (product), the relative competitiveness (price) and the distribution channel (place). The promotion part, due to its relevance to the current modelling exercise, is forming a separate group, called marketing activity (group 3). As detailed in the scope (Chapter 1.2) current research is focusing on the predictive campaign cost modelling of current account, credit card and savings accounts however all metrics in this group should be broken down by all products. The qualitative research highlighted that cross-sale[98] activity can account for a portion of new and existing customer recruitment and certainly can increase the share of wallet or even loyalty. The following variable categories have been identified:

- ***Product Openings by channel:*** This category should contain the number of products opened by new to group and existing customers split by channel (branch, telephony distribution, internet) for all products (current accounts, credit cards, savings products, unsecured personal loans, mortgages and general insurance products). Each split should initially contain the number of all products opened by product type to summarise the different splits. This detailed split will allow the analyst to understand the impact of cross-sale activity and the direction

[96] For more information please visit www.gfk.com
[97] Product, Price, Promotion, Place [KOTLER-KELLER 2012, p.25]
[98] Cross-sale is when the customer purchases two product at the same time or within a short space of time

with the strength of the relationship between products. Each product category will result at least seven variables therefore a review will be required before modelling. These numerical variables are generally available in the internal account system of the financial institution.

- ***Market share by product category and channel:*** Similarly to the product openings, the market share split by product and channel has been identified as not just target variable but explanatory variable too. As one of the interviewed industry professional explained, the existing market penetration to other products (other than the targeted banking, savings and credit card) should account for a portion of new customer recruitment. It is even more pronounced when the stock market share figures are used as predictor variables. It was also suggested that certain channels such as online and agencies may explain more than branches and telephony distribution (TD) from the variance of the target variable as these channels are highly competitive. For example, agents or online financial aggregator sites can offer multiple products at the same time therefore a high market share stock position highlights the existence of a competitive product and the popularity, which itself can increase sales. The market stock and new business (which is always related to the actual campaign) share volumes as percentage split channel (branch, telephony distribution, internet) for all products (current accounts, credit cards, savings products, unsecured personal loans, mortgages and general insurance products) are required at this stage as well as the total market share by product and channel. This information is generally available for financial organisations through external market research agencies. The evaluation period is monthly and the data is split by total market and the data requestor bank. In some cases there is an option to split the total market into a peer set and the rest of the market. In current research the simulation data is based on market share information the bank subscribed for however the actual figures have been desensitised to only reflect the extent and direction of the change between two data point making the data useful for the modelling purpose.

- ***Market competitiveness***: Financial products do not normally have upfront price therefore the evaluation of the price or the cost for the customers is sometimes not straightforward. Due to the special need for financial products (for example current account for most salary payments or to complete certain transactions that are required to be paid via bank transfer, cheques, etc.) customers not always can delay their financial product purchase (unlike with other products, like travel or jewellery, etc.) therefore the price (cost, interest, etc.) will not drive the purchase decision but will primarily influence the provider selection. The cost or benefit of financial products (e.g. APR or AER) are normally linked to the base rate of the central bank across all providers with usually a small[99] variance therefore the best measure for the effect of price is measuring the relative position of the product proposition in the market. Using the APR information for credit cards and the AER information for savings accounts as well as the fees vs. rewards for the flagship current account proposition, the competitiveness of these products can be measured by three separate variables. Another set of variables within the market competitiveness is the official product comparison in the media. This has two main sources in the United Kingdom: the first one is the so called 'best buy tables'[100], which helps customer to review current market offerings and make better financial decisions. The

[99] Based on current market condition it is estimate to be with a +/- range of 7 percentage points.
[100] For more information on best buy tables please see http://moneyfacts.co.uk/.

interviewed industry experts agreed that a top three position and the number of mentions of the branded product in savings, current account and credit card categories on the best buy table have directly proportional linear impact on purchase probability and therefore new customer recruitment. The second source is the analysis of industry reviews by experts of financial aggregator (or comparison) sites[101]. For example, one of the interviewed experts detailed an example where a positive review of Martin Lewis (money saving expert) resulted over 40% uplift in current account sales for over 6 weeks in 2013 therefore the number of mentions by industry experts is added as a potential predictor.

Group 3 – Marketing Activity: The third group of variables are grouped around the marketing activity and marketing cost. This category covers the key marketing events, the brand power and the various cuts of the resource spent on marketing activity. This section is important and require accurate and detailed variables for a precise campaign cost estimation model. The following key variable categories should be assessed:

- ***Marketing events:*** The marketing activity can be split into many levels based on the responses of the industry professionals. There is a basic or 'unconscious' level when purely the size of the organisation (through the number of branded branches, etc.) with its products and employees attract a certain number of customers without any other specific marketing activity. The next level is the brand building where the organisation is consciously reaching out for potential costumers through general messages and advertisement about the organisation. The third level identified by the industry professionals was the campaign activity when the financial organisation promotes its products or services to achieve certain sales level. The highest level of the marketing activity is the one-to-one marketing when organisations are using complex algorithms to determine the customers' needs individually and even offer tailored product propositions. Current research is focusing on the third level to identify the optimal cost and to forecast the success of future campaigns. For this, the key campaign activities of the financial institutions such as sponsorship events (sponsored sport, cultural, educational events); direct marketing events (direct post and e-mailing campaigns) and media campaigns (TV, radio, press, etc.) should be identified.

- ***Brand awareness:*** Another angle of the new customer recruitment capability of the financial institution is the brand power. It is expected by the industry professionals that the unaided brand awareness variable combined with the brand warmth would provide some indication of the success of a new campaign.

- ***Media spend by product type:*** The following two groups of variables are detailing the monetary resource spend on the campaign activity. This provides the basis for the cost prediction. It is important to capture all marketing spend on campaign and brand awareness activity and not just the specific product related ones as the interrelations and the marginal changes caused by cross-campaign activity should also be captured and factored into the model to increase prediction accuracy. It was suggested by the industry professionals to collect all marketing spend for the most granular time-period available (e.g. days or weeks), then proportionate the spend to the marketing activity instead of using the invoice payment

[101] For example Martin Lewis's www.moneysavingexpert.com

date. With other words, every cost should be spread throughout the campaign time reflecting highs and lows in the campaign activity. For example, if we assume that a four weeks TV campaign would cost £40,000 with one airing per week then the cost should be split into four equal parts and each part should be assigned to the weeks the TV advert was aired. Theoretically this way, the time between the marketing spend (i.e., the actual market activity) and the customer recruitment would be synchronised and the amount would be a true reflection of the campaign size. This type of data split also enables post modelling analysis to determine the required length of a campaign for certain products or media type. It has been suggested by industry professionals that not just the assessed organisation's own media spend but also the relative media spend with the total market media spend should be collected to understand the positive and negative externality[102]. The following variables should be compiled: marketing spend by total market, the assessed bank and the spend share for total spend, brand building and services, business banking products and services, general insurance, investments, current accounts, mortgages, unsecured personal loans, credit cards and savings products.

- ***Media spend by media type:*** Following the above detailed product centric split of marketing cost the same spend should be split by media type too. It will enable the analyst to determine the most effective media type and the average conversion rate by type for strategic focus alignment. The total marketing spend should be split by total market, the assessed bank and the spend share for spend on cinema, digital, direct mail, outdoor (such as shop window banners, and other banners for outdoor events), press, radio and TV advertisement elements of the campaign. Given that this information is required for the entire market split by competitors (or key peers) and for the organisation it is suggested to purchase the data from an external market research agency.[103] These agencies collect data 24 hours a day for 7 days a week and based on an extensive database of cost associated to certain marketing activities they collect, categorise and estimate money spend on each media activity by product, media, brand, etc. For example, for press activity all key newspapers are reviewed and based on the day of the advertisement (which influences the amount of copies sold), the page and size a cost is estimated for that day for the advertiser institution split by product. All these spends then collected and aggregated up. Current research is using desensitised simulation data derived and altered from data used at the selected retail bank from a market research agency.

Group 4 – External Factors: The final group of variables is centred around external factors and events that can impact the effect of other variables as well as describe the economic environment and partially explain the variance of the target variable.

- ***Economic environment:*** The economic environment can influence bank product offerings and customer needs in many ways. For example, in a low interest rate environment savers are encouraged to invest rather than to keep funds in instant savings and bond products or higher

[102] Competitor marketing activity can be pro- and counter active from the organisation's perspective. Although most often it is counter active, there are examples for increased sales at other competitors. For example, when a new product advert highlights the latent need of the customer for a service or product purchase but the customer is either loyal to a financial institution or will shop around instead of following the brand the marketing campaign was set up by.
[103] It is also advised to analyse the accuracy of data provided by the external agency through the comparison of internal spend data and the relevant agency data.

unemployment rate can forecast an increased need for unsecured personal loans and credit cards. When considering the economic environment current research includes the inflation rate, average weekly gold price, GDP, consumer confidence index[104], business confidence index[105], average weekly earnings, unemployment rate, number of employed men and women (which can be a good indication of family increase change), gross disposable income, median house prices, SVR[106] mortgage rates, Bond rates and base rate.

- ***Weather:*** In social and behavioural sciences, the effect of weather and ambient temperature on consumer preferences is an increasingly popular topic. For example, recent researches suggested that warm (vs. cool) temperatures dispose consumers towards others' opinions as the basis for product preferences, stock price forecasts and betting[107]. [HUANG et al., 2014] On simple terms, taking weather into consideration means the analyst is trying to answer certain anomalies in the variance of the target variable. The underlying assumption is that in specific weather conditions (e.g. at heatwave or unusually low temperature or major storms followed by floods) can delay the marketing campaigns' impact on product opening, which effect should be isolated and excluded from the explanatory power of the cost elements. For this purpose, current research is using the average temperature (in Celsius degrees) and the average rainfall (UK average in mm) as explanatory variables.

- ***Seasonality:*** Seasonality has been identified as an important predictor by both the industry professionals and the literature. Industry professionals highlights the tax year end (or the so called cross-tax year) period when the interest free individual savings allowance (ISA) is increased by the government every year. This event naturally creates a high demand period or season for ISA products. Other examples detail the deterministic seasonality in mortgage transfers and house prices [NGAI-TENREYRO, 2014] or the complexity of modelling seasonality for credit cards and other services in general [SWARTZ-IACOBUCCI, 2000]. Current research is focusing on three seasonal flag through binary variables (which determine whether the campaign period is on or off season): ISA season (for savings products), Student Season (for current accounts) and Savings bond or fixed term savings products (for savings products). Credit cards sales is expected to be even throughout the year however there are peaks before religious and public holidays (such as Christmas, Diwali, Easter, etc.), which variable is covered by the next ("events") category.

[104] The Consumer Confidence Index is (CCI) is a survey based index, measuring the health of the economy from the perspective of the consumer. "*The index is based on consumers' perceptions of current business and employment conditions, as well as their expectations for six months hence regarding business conditions, employment, and income*." for more information please visit: https://www.conference-board.org/pdf_free/press/TechnicalPDF_4134_1298367128.pdf

[105]*The Business Confidence Index (BCI) is based on enterprises' assessment of production, orders and stocks as well as its current position and expectations for the immediate future. Opinions compared to a "normal" state are collected and the difference between positive and negative answers provides a qualitative index on economic conditions."* [OECD, 2016]

[106] SVR or standard variable rate is a type of mortgage rate. This rate is normally the 'default' rate without any limited-term deals or discounts attached. This represents the highest mortgage rate in general.

[107] Furthermore, warm temperatures increased the participants' perception of social closeness to other decision-makers thus leading them to consider the opinions of those decision-makers to have greater validity. This increased validity, in turn, rendered them more likely to conform to the crowd.

- ***Events:*** The final external variable category is the group of events. These events are linked to general increase in demand (for example personal loans prior to Christmas), which naturally drives more customers to financial providers and are also linked to events, which may be related to previous sponsorship activities or just general uncertainty. The Olympic games are generally good examples where a sportsman/woman is (or used to be) the brand ambassador of a financial provider. The further they go in the competition the more airtime they get, which results stronger brand association with the bank. This can eventually be translated to new customer recruitment. Another example is political uncertainty. Since Brexit "*both house demand and supply indicators remained somewhat weaker than in 2015 and early 2016*."[108] It is suspected by industry professionals that the economic environment is contributing to the slowed housing market as people wait for more clarity on the future before making major decisions such as buying or selling properties. In order to take events into consideration, current research includes the number of working days per period, a binary variable for all national and international sport events broadcasted by television [109] where the UK is participating and another binary variable for key national or international political events such as elections, referendums and financial industry related scandals.

The next step within stage two is the **setting of time horizon** for the campaign cost prediction model. This is required for data extraction and the for the model. Campaign cost modelling is a business related econometric problem and as such, the data collected and analysed in relation to the model is going to be a time series data. The analyst and researcher have the following two options:

1. The first option is to isolate the different campaign periods and use aggregated data on campaign level. This approach treats the different campaigns as (almost) separate observations which will reduce the impact of autoregressive errors and provides a data format and structure aligned to the basic assumptions of the multivariate linear regression method. Although this way the analyst will exclude the interrelations of different campaigns, he/she will not factor in the impact of some soft metrics such as brand awareness, which may not only be linked to campaign activity and will not take into consideration the effect of seasonality. The sample size, with other words the number of campaigns in this context, can also cause an issue as a minimum of 30-50 campaigns would be required for the cost model.

2. The second option is to use continuous time series data covering an adequately[110] large period of time and set the unit of time according to the required granularity. The benefit of this method is that the model can potentially be built on a larger sample size. Furthermore, factors like the 'natural' customer recruitment rate, the amount of new customers purchasing product without any specific campaign activity can be measured. Although time series data can violate the basic assumptions of the linear regression modelling methodology (through autocorrelated errors for example), there are corrective actions and preventive measures the analyst can take to make the time series data usable for linear regression. When considering this option the analyst or researcher should also make a decision about data granularity. While making the decision, annual aggregation would potentially eliminate the seasonality caused dependency

[108] Source: https://www.gov.uk/government/publications/uk-house-price-index-summary-september-2016/uk-house-price-index-summary-september-2016
[109] On BBC, ITV or SKY NEWS
[110] In order to decide what is adequate the analyst should consider the minimum sample size and the number of years of data still relevant to given research (for example there are no major changes in the political or economic environment).

in the error terms, it also reduces the sample size significantly. The optimal solution is somewhere between daily, weekly or monthly level.

Although both methods have their rationales, author has chosen the second option since the data extraction is followed by the current step and therefore the amount of data extracted is in control of the author and not based on survey responses or on a secondary modelling database. Furthermore, author sees the second option more accurate for cost modelling due to the availability of the large base dataset. As the last major changes in the economic, political and legislative environment happened over four years ago (2010-2012[111]), the aim is to collect the data from the past four years. Based on Table 9 the advised data granularity is weekly, which would provide the linear regression model building exercise with just over 200 data points.

The last step of the current stage is the **data extraction**. There are two main sources identified by the author. The primary source is the simulation database extracted from the participating financial organisation. The simulation database is based on data directly extracted form the bank's own Oracle and Hadoop databases using SAS, Impala and Business Object database management tools. The simulation database also incorporates desensitised information provided to the financial organisation by third party marketing agencies. The primary source covers the first three groups of the modelling database. The last group of variables is coming from publicly available sources, such as Office for National Statistics, Government services and information and Bank of England. The following table summarises the different sources directly (and indirectly) applied for the simulation database creation:

Table 16: The source of predictor variables

Level 3	Source
• Staff, Network size, • Branch Appointments, Telephone Distribution Appointments, Website visits, • Product Openings (Current Account, Credit Card, General Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans) • Internal Customer Satisfaction survey • Sponsorship Event, Direct Mailing Campaign, TV campaign	Internal Accounting Database Internal Marketing Database Internal Business Intelligence Database
• External Customer Satisfaction rate • Market Share Stock and New Business Market Share (Current Account, Credit Card, Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans) • Savings rate competitiveness, number of mentions in press • Internal and Market marketing spend by product (Total, Current Account, Credit Card, Insurance, Investment, Mortgage, Savings, Unsecured Personal Loans) • Internal and Market marketing spend by media type (Cinema, Digital, Direct Mail, Outdoor, Press, Radio, TV) • Brand awareness, Brand warmth	Internal Management Information System from Market Research Agency
• Unemployment rate, Inflation, GDP	www.ons.co.uk
• Median house price, Base rate, Avg. Bond rate, Avg. Mortgage rate, Gold Price	www.bankofengland.co.uk
• CCI, BCI	data.oecd.org
• Average Temperature, Average rainfall	www.metoffice.gov.uk
• ISA Season, Student season, Bond season, working day counter	www.gov.co.uk
• Sport Event flag, Political Event flag	Various newspaper articles

Source: Compiled by author

[111] Financial Services Bill in January 2012 to create new regulatory bodies such as Financial Policy Committee (monitoring risks of the financial sector and financial regulations), Prudential Regulation Authority (supervising soundness of financial organisations) and Financial Conduct Authority (protect consumers)

During the data extraction author identified certain datasets, which are not available on the selected granularity (weekly) therefore the aggregation (for daily data), phasing (for monthly or quarterly information) or calculation of the end of period stock position was required. Due to the space limitations of current dissertation, the detailed list of variables and their methodologies on a variable by variable basis can be viewed in the appendix (A5.). In some cases, further transformation of the data or a qualitative measure was required and author has created a set of Bernoulli and proxy variables during the initial data collection process. The following **binary variables** have been created: ISA, Student account and Bond Season flags (the value is 1 if at least one day of the week is part of the respective seasons and 0 if none of the working days are part of the seasonal periods).

The following **proxy variables** have been set up: Service level competitiveness (1-4 reverse ranking based on the weekly Customer Satisfaction lead table where 4 represents the first position and 1 is the 4th or more rank); Market competitiveness of various savings products (0-4 ranking based on AER rates by providers within the same[112] product category. These ranks are in reverse order where highest=4; 2nd=3; 3rd=2; 4th or more=1, not in competition=0); Best Buy Mention Ranks (0-4 ranking based on the total number of best buy table mentions in any given week. Similarly to previous categories, the ranks are in reverse order where the highest=4; 2nd=3; 3rd=2; 4th or more=1, no mentions=0); TV Campaign indicator (total number of active TV campaigns at any given week, a TV campaign is considered active if it the TV advertisement of the campaign is aired for at least one weekday of the week); Working days (total number of working days where days between Monday and Friday are considered 1, Saturday as 0.5, Sunday and official Bank Holidays in England as 0).

Author decided to create the ranking variables in reverse order of the actual rank. The underlying logic is that the higher the rank, the more competitive the provider is and higher competitiveness should be marked by higher value. Author used equally distributed integers for the competitiveness ranking as the distance between two providers in any ranking category is not assessed by the current model. Furthermore, current study limits (or stated otherwise) the number of competitive options as any rank equal to or higher than the 4th are considered as the same level of competitiveness. For the current modelling exercise it is assumed that there is only a marginal difference in competitiveness above the 4th rank and therefore it is not assessed.

The data extraction is followed by the data formatting. Table 17 demonstrates the formatting and naming principles through an actual section of the final pre-modelling simulation dataset. The entire dataset is on weekly bases and in ascending order by week. A new column has been added to mark

Table 17: Preview of the pre-modelling simulation dataset

PER	WEEK	GDP	INFLATION	CCI	BCI	EMPLOYMENT	…
1	27/02/2012	424211	3.4	97.9447	100.4282	70.5	…
2	05/03/2012	424834	3.5	97.9116	100.4702	70.5	…
3	12/03/2012	424712	3.0	97.9891	100.3368	70.7	…
4	19/03/2012	424590	2.9	98.1788	100.0025	70.9	…
5	26/03/2012	424468	2.4	98.3726	99.9295	71.1	…
6	02/04/2012	426085	2.6	98.4382	100.0123	71.2	…
7	09/04/2012	427702	2.5	98.4418	99.9652	71.1	…
…	…	…	…	…	…	…	…

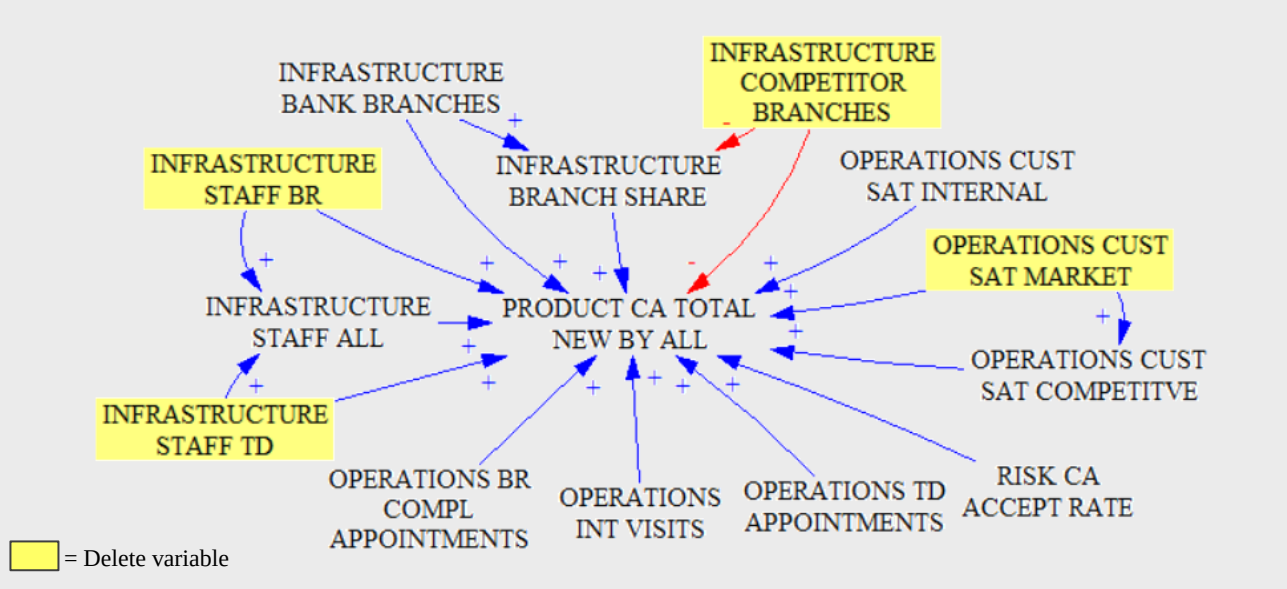
Source: Compiled by author

[112] Same product type (ISA, Instant Access, Bond, etc.) and same bonus term (1 year, 2 years, 3 years, etc.)

the number of the periods, with other words this is the incrementing number of weeks since the first data point. The naming convention of variables follows the one outlined in appendix 5. The final data mart of this stage contains 185 variables and 205 observations (data of consecutive weeks).

4.2.3 Stage III. – Pre-modelling analysis

The first step of the pre-modelling analysis is the assessment and review of all identified variables. The aim is to explore the relationship between the suspected predictors and ultimately exclude the ones where there is a functional relationship. Figure 22 details the review of Group 1 variables.



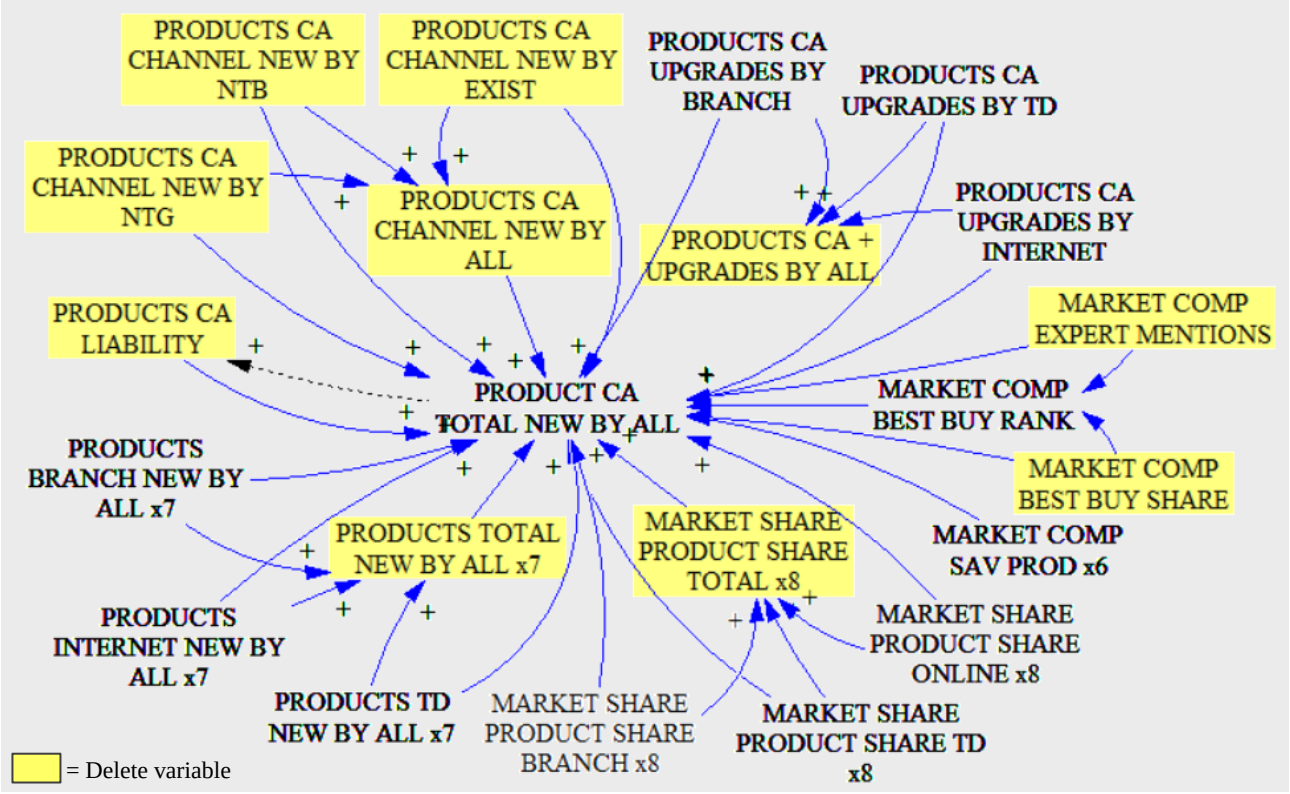
Source: Compiled by author

Figure 22: Initial assessment of variable interrelations for Group 1

The highlighted variables have been excluded from the current model as the effect of these variables have been represented by other predictors through aggregation or a function. For example, the total volume of staff is the sum of the channel specific staff volumes. Although a channel specific split of the adviser structure can provide decision makers with valuable insight if the analyst is focusing on the impact of different advisor groups, for current cost modelling exercise author excluded the branch and telephony distribution specific staff volumes. The total volume of competitor branches predictor has also been excluded as it is expected that the market share of branch footprint has greater explanatory power compared to the market volume only. The last item to be excluded from group 1 was the absolute external customer satisfaction score as it was replaced by the measure relative to competitors.

Figure 23 shows the assessment of group 2 variables. Due to the same variable pattern across different products author simplified the assessment by showing main category groups and highlighting the number instances by ‘x7’ or ‘x8’. Most variables related to current account (CA) sales, such as openings via different channels and by customer types, have been excluded as the target variable is the aggregate of these predictors. Furthermore, the liability (the total value of deposits) of current accounts and the CA market shares are partially explained by the target variable. Therefore, these predictors cannot form part of the further analysis. Other aggregate variables such as total product sales in other product groups made by Bank A or the Market, have been excluded as well as the total

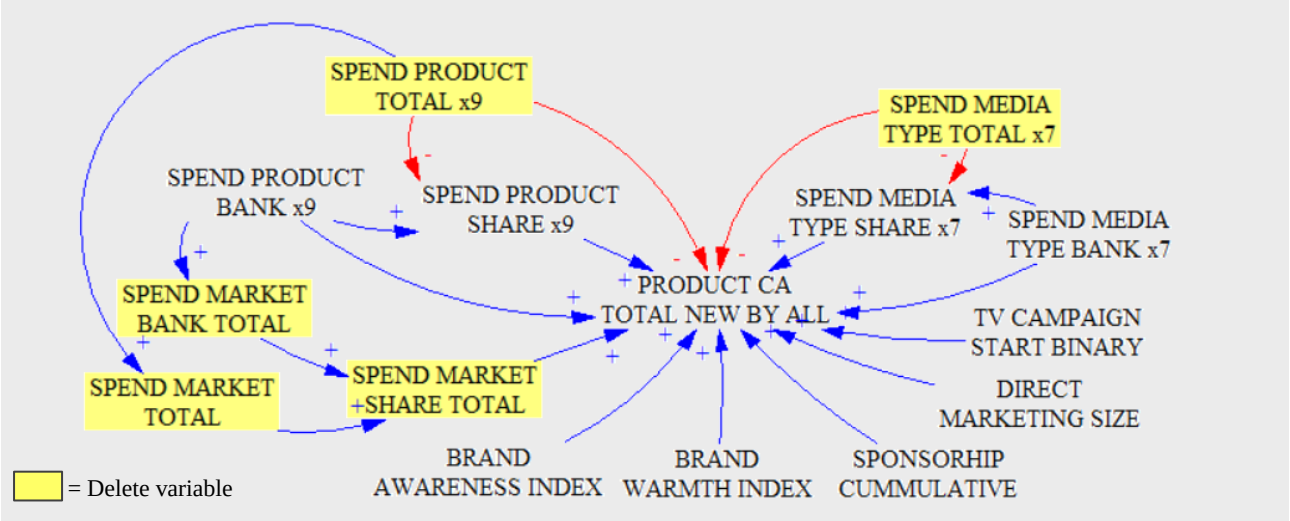
number of mentions in the Best Buy tables because the Best Buy table shares are represented better by the Best Buy Rank variable.



Source: Compiled by author

Figure 23: Initial assessment of variable interrelations for Group 2

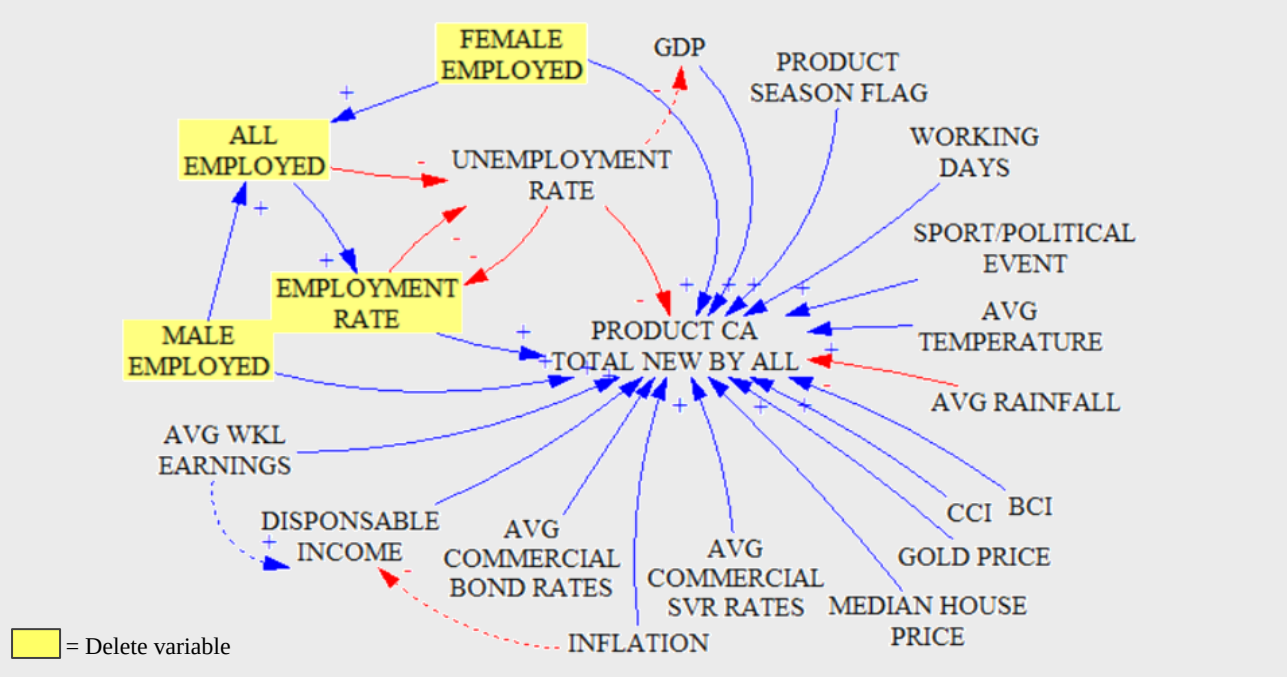
Figure 24 demonstrates the variables centred around the marketing spend and marketing activity. Variables related to the absolute value of total market media spend by product or media type have been excluded as the relevant Bank A spend share metrics contain more information compared to the total market value only. The total media spend variables related to both the Bank A and the market with the total market share figure have been excluded as these variables, are the aggregation of separate spend categories by product and media type.



Source: Compiled by author

Figure 24: Initial assessment of variable interrelations for Group 3

The final assessment chart is describing the relationship between the external factors. Most of these variables are interlinked. In some cases there is an indirect link between two variables but most of them are influenced by the same social-economic events (for example the financial crisis).



Source: Compiled by author

Figure 25: Initial assessment of variable interrelations for Group 4

Having noted that these variables have a potential underlying factor structure, author only removes the different components of the unemployment rate metric (total volume of employed, split by gender, employment rate) and will assess the correlation between the variables in case these variables should be aggregated to factors using the dimension reduction method.

The second step of the pre-modelling analysis is the assessment of the correlation matrix. For this author is using the Pearson correlation matrix on all remaining 121 variables[113] and a simple t test (p value) for each of the correlation pairs to determine whether there is evidence for the significance of the correlation.[114] Please find the Detailed Correlation Matrix in the appendix 6. Based on the initial analysis of the correlation matrix author has identified 26 variables with medium to high correlation[115] with at least 20% of all remaining variables. Following the logical analysis of these predictors, author firstly excluded the Branch infrastructure related predictors as these show medium to high correlation with over 50% of the rest of the variables. The absolute value of the branch numbers across the assessed period is decreasing by an approximate rate of 1% every quarter. The relative branch share is almost constant for the assessed 205 weeks, which suggests a similar branch closure trend across the market. Therefore, there is no logical evidence for the distinctive impact of the branch volumes on Bank A's product openings compared to the peer group. Secondly, during the assessment the current account market share related predictors have been excluded due to the high correlations. Furthermore, these variables are partially explained by the target variable. In case of two predictor groups (Unsecured Personal Loan openings split by the channel of opening and the Current Account Product Upgrade to the assessed Flagship product split by the channel of upgrade) both show high

[113] Remaining after the logical predictor relationship assessment and simplification detailed above.
[114] Correlation between two variables is not considered if the p>0,05.
[115] For thresholds please see the Correlation section of Chapter 3.3.3 Stage III. – Pre-modelling analysis

correlation with the target variable split by channel and the different channel values of the respective variable groups therefore the channel split was replaced by the total vales (for example UPL Openings split by channel with the total UPL Openings of Bank A). The remaining 11 variables have shown moderate to high correlation with over 50% of all variables but as these predictors are all linked to the macroeconomic environment the presence of these variables is vital and a reduction method is required to keep the meaning but cease the unwanted correlation. Following stage three, the final dataset before the dimension reduction exercise contains 108 variables.

4.2.4 Stage IV. – Reduction of dimensions

The pre-modelling analysis highlighted a group of variables[116] with medium to high correlation with over 50% of all predictors. The logical consideration suggests a latent factor structure behind these variables centred around the general health of the economy, employment and the interest rate environment. If the factor structure exist, it can simplify the variable system and reduce the risk of multicollinearity. Author at current stage will perform a Common Factor Analysis (CFA) to test the suggestion for the presence of latent factor structure among the macroeconomic variables and to explore the interrelations of all remaining variables.

The first step of the factor analysis is the **assessment of the prerequisites**. Following the logical consideration (above) the analysis continues with the statistical assessment. Although the analysis is primarily testing the 11 macroeconomics related metrics, a second factor analysis will be conducted where all 108 remaining variables are included for a wider assessment of the interrelations between the predictor variables. The following listing details the first CFA model only but will also overlay the results of the second analysis.

a) All variables included into the factor analysis are metric variables.
b) The correlation matrix has already been created as part of the pre-modelling analysis stage. As detailed previously, the analysis of the correlation matrix (Appendix 6) suggests the existence of a latent factor structure. Based on the correlations and logical assessment, a two or three factors structure is expected behind the 11 corresponding variables however not all variables can be assigned to a single factor.
c) The further visual assessment (using scatter plots) of the relationship between these variables (Appendix 7) highlighted that the Business Confidence Index does not meet the linearity criteria therefore it is excluded from the analysis and the multivariate regression model.
d) The next step is to statistically examine if correlations between the variables are not due to randomness. As Table 18 shows, we can reject the null hypothesis (there are no common factors between the initial variables) of Bartlett's test for sphericity.

Table 18: Bartlett's test for sphericity

Significance Test Based on 205 Observations			
Test	**DF**	**Chi-Square**	**Pr> ChiSq**
H0: No common factors	45	5069.3092	< 0.0001
HA: At least common factor			

Source: Compiled by author based on SAS output (please find screen-print in Appendix 8)

[116] Variables in the group: Gold Price, Average Commercial SVR rate, Average Commercial Bond Rate, Average Weekly Earnings, Employment Rate, Inflation, Consumer Confidence Index, Business Confidence Index, Median House Prices, GDI, GDP. For more information please see Appendix 6.

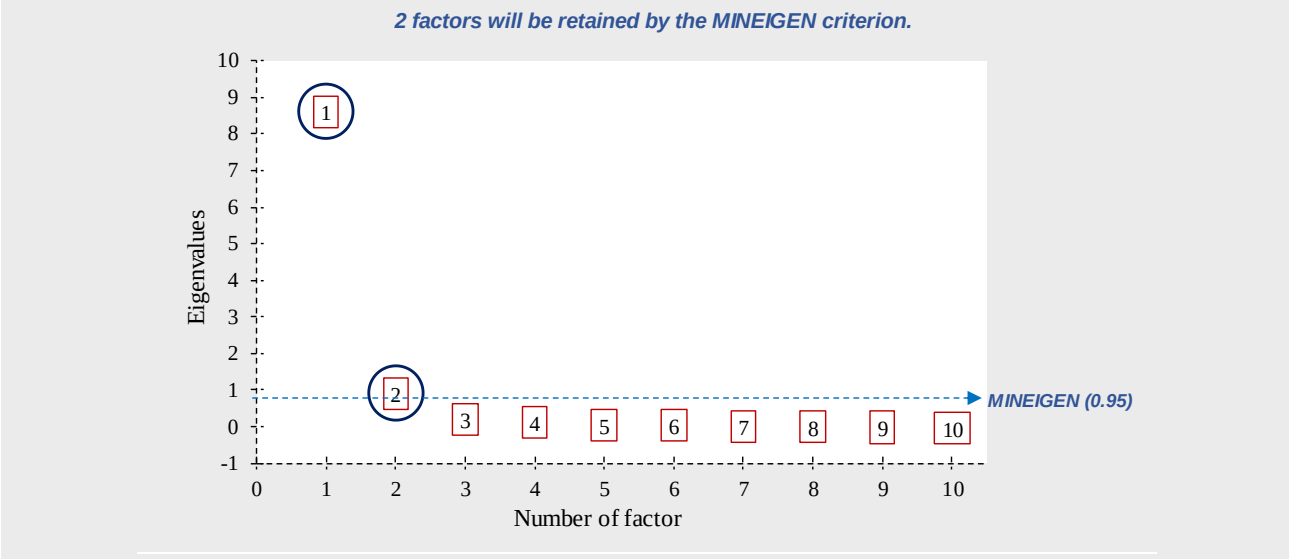
e) The assessment of the anti-image matrix confirms that all variables are adequate for the factor analysis as the MSA (Measure of Sampling Adequacy) is always above 0.5 (the MSA values are ranging between 70.78 and 91.71).
f) The final step is the test for sampling adequacy applying the Kaiser-Meyer-Olkin criterion. The overall MSA is 0.87, which is greater than the 0.5 cut-off level and even higher than the stricter 0.7 value.
g) Table 19 shows the result of the Sample Adequacy test.

Table 19: Kaiser's Measure of Sampling Adequacy

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.86392821			
Variable	**MSA**	**Variable**	**MSA**
GOLD_PRICE	**0.91**	INFLATION	**0.87**
AVG_COMMERCIAL_SVR_RATES	**0.83**	CCI	**0.80**
AVG_COMMERCIAL_BOND_RATES	**0.81**	MEDIAN_HOUSE_PRICES	**0.81**
AVG_WKL_EARNINGS	**0.92**	GDI (Gross Disposable Income)	**0.91**
EMPLOYMENT_RATE	**0.92**	GDP	**0.87**

Source: Compiled by author based on SAS output (please find screen-print in Appendix 8)

Following the statistical confirmation of data adequacy for factor analysis the next task is to **create the initial factor extraction and determine the number of factors to retain**. Author's aim with current dimension reduction exercise is to determine the minimum number of factors that explain sufficiently[117] the common variance of the macroeconomic variables. As this aim can be achieved by any factor extraction method, author is applying the default 'principal factor' factoring method of the SAS statistical software package. Practically it means that the initial number of factors is equal to the number of variables are included (similarly to the Principal Component method) then by using a combination of the selection criteria outlined in chapter 3.3.4 (the visual scree-test, assessment of communalities and checking the interpretability) the final factors are selected and interpreted. Based on the Scree plot of the initial factor extraction, shown on Figure 26, the analyst can either keep one or two factors. The MINEIGEN criterion, which is based on the average prior communalities of the variables (0.913 – Appendix 8) suggests two factors however the second eigenvalue is less than one.



Source: Compiled by author based on SAS output (please find screen-print in Appendix 8)
Figure 26: Scree Plot Eigenvalues (macroeconomic variables)

[117] Explain at least 80% of the variance.

The first factor is accountable for 87.5% of the total variance with an eigenvalue of 8.60. The second variable explains 9.3% of the total variance with an eigenvalue of 0.91[118]. Although the inclusion of only one factor would result the minimum number of possible factors when the unrotated factor structure is assessed, the originally suspected 2-factor latent structure emerges therefore author decided to include the first two variables.

The next step is the **rotation of the factors** in order make the interpretation of the factors easier and the **analysis of factor loadings**. For current cost modelling exercise author is using the varimax orthogonal transformation to reduce the risk of multicollinearity (with this method the different dimensions or axis are perpendicular to each other, which means the factors are uncorrelated).[119] As stated earlier author considers a variable loading on a factor if the factor loading is greater than or equal to |0.50|. Table 20 shows the result of the orthogonal factor rotation. Values with high factor loading (greater than or equal to |0.70|) are marked with an asterisk (*). The analysis of factor loadings highlights that the Average Commercial SVR variable is loading on both factors and cannot be added to any of the factors. Furthermore, Factor 1 covers the Average Weekly Earnings, Employment Rate, Inflation, Median House prices, Gross Disposable Income and the GDP metrics. The GDP variable has the highest factor loading on Factor 2 however it still below the cut-off level set by current study. Factor 2 is defined by the Gold Price, Bond Rate and the Consumer Confidence Index.

Table 20: Rotated factor pattern

Values greater than 0.7 are flagged by an *				
Test	**Factor 1**		**Factor 2**	
GOLD_PRICE	-0.38		-0.80	*
AVG_COMMERCIAL_SVR_RATES	0.61		0.68	
AVG_COMMERCIAL_BOND_RATES	-0.30		-0.88	*
AVG_WKL_EARNINGS	0.87	*	0.43	
EMPLOYMENT_RATE	0.78	*	0.44	
INFLATION	-0.88	*	-0.41	
CCI	0.34		0.87	*
MEDIAN_HOUSE_PRICES	0.89	*	0.35	
GDI (Gross Disposable Income)	0.75	*	0.35	
GDP	0.80	*	0.49	

Source: Compiled by author based on SAS output (please find screen-print in Appendix 8)

The next step is **naming the factors,** which requires a certain level of creativity from the analyst. Factor 1 is linked to employment rate, weekly net and gross earnings, productivity on a domestic level and house prices. Author believes these variables are mainly related to the effective demand in the economy. The fact that higher gross and net weekly earnings or the GDP result higher factor value supports the observation. Furthermore, the factor is also linked to the house prices, which can similarly be connected to the effective demand (as domestic house prices increase faster if there is a high demand by buyers with an effective purchasing capability). At the same time, Factor 2 is centred around economic stability. The factor is directly proportional with the Consumer Confidence Index. When the index is high, customers are likely to spend their savings, they are positive about the economic outlook and have confidence in continuous income stream. When the index is low, consumers delay investments and major purchases (such as new car, house, expensive holiday, etc.).

[118] The complete SAS output of Eigenvalues of the Reduced Correlation Matrix is available in Appendix 8.
[119] For more detail please see section 3 within Chapter 3.3.4.

During economic hardship, the index is low[120] and consumers are turning their attention to safe means of investment, such as fixed rate bonds and gold, to preserve the value of their savings. Factor 2 is indirectly proportional to these variables, which means higher bond rates and gold price lower the value of the factor. Based on the analysis above the following factors have been created:

- Factor 1: Effective Demand *(explains 87.5% of the common variance)*
- Factor 2: Economic Stability *(explains 9.3% of the common variance)*

The last step of the factor analysis is **the extraction of factor scores** and the replacement of the original variables. The two newly created factors are calculated and outputted by the statistical software package[121] and the analyst is only required to replace the variables in the modelling datamart by the new factor variables. Author has also conducted a wider, exploratory factor analysis using the same methodology outlined above, to understand if there is a latent factor structure behind the remaining variables. The analysis highlighted that there is no such variables structure however the sales variables clustered around loan products such as mortgages, unsecured personal loans and credit cards have a moderate interlink. At this stage, author is including the mentioned variables into the regression analysis without any further dimension reduction. Furthermore, author does not see the need for a principal component analysis. Following stage four, the final dataset for the multivariate linear regression model contains 97 original predictors and 2 factor variables.

4.2.5 Stage V. – Multivariate Linear regression modelling

The first step of the multivariate linear regression modelling is the **selection of the model building strategy**. The model is intended for an accurate cost prediction therefore the author has chosen to include all (meaningful) variables into the initial model from the modelling datamart. The number of variables will eventually be reduced using a variable optimisation method and logical consideration.

The second step is the **definition of the full and challenger models** (with a variable optimisation technique). The author has established five models (including the initial full model) for the hypothesis testing. The following section summarises the models and their analytical rational:

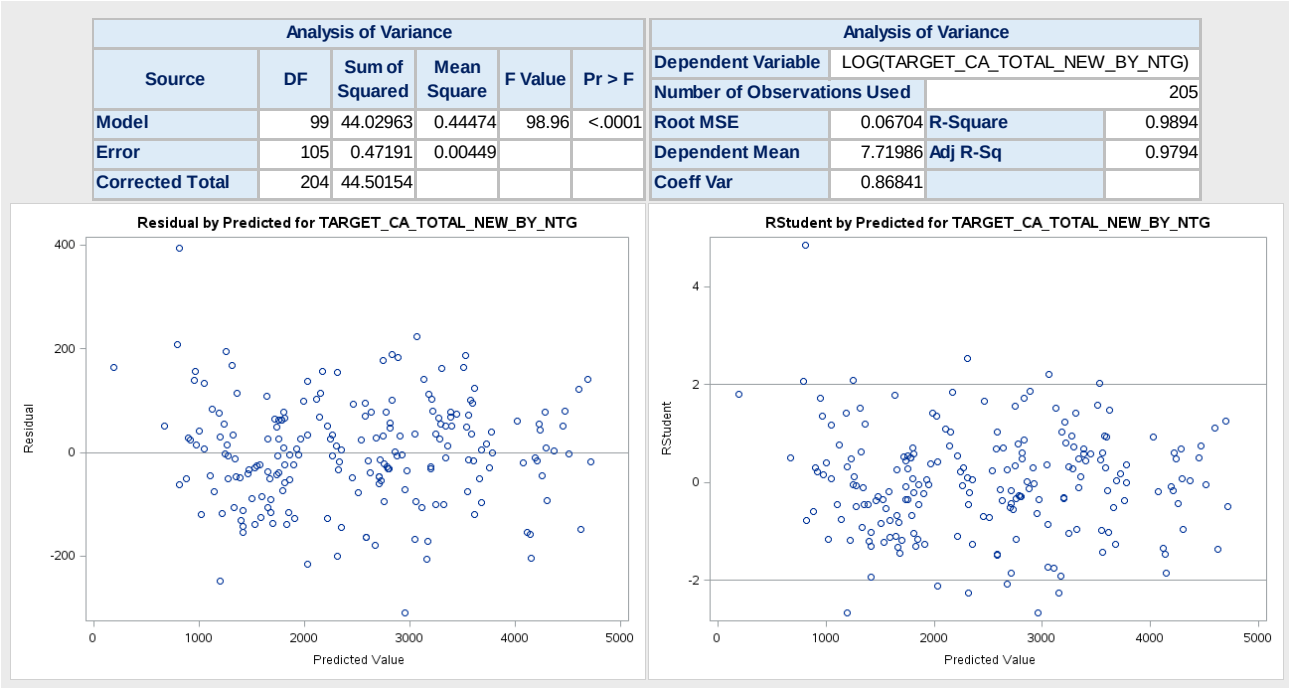
- Full Model: The full model includes all initial (99) variables from the Stage IV. Datamart. This model will form the basis of the comparison. This model is expected to explain the highest proportion of variance however the high number of variables can make the model inefficient for analytical and cost predictive purposes. Author keeps in mind that the result of this modelling exercise must be practical enough for finance professionals to adapt in daily planning activity therefore a reduced model is required.
- Model 1: For the first challenger model author applies Stepwise selection from all data mart variables. This is the auto-rationalised version of the full model without any other logical consideration. The stepwise selection criteria is detailed in the completion of the challenger models section of current chapter.
- Model 2: The next model also applies stepwise selection method but, as a further manual variable reduction method concens, author replaces the existing channel split variables with

[120] Based on the data available on https://data.oecd.org/leadind/consumer-confidence-index-cci.htm one can see that local minimums of the graph are in 1973 (First oil crisis), 1979 (Second oil crisis), 1994 (Mexican Peso Crisis), 2002 (Dotcom Crash), 2009 (Financial Crunch) which all relate to world economic crisis and instabilities.
[121] In SAS EG 5.1 it is available through the "NFACTORS= SCORES OUT= ;" argument of the Proc Factor procedure.

their product level aggregates to complete an assessment excluding the channel impact. This way the model will start with 83 variables instead of the 99 from the Stage IV. Datamart.

- Model 3: The third model, similarly to the previous challengers, uses stepwise selection but excludes all market share variables and only includes the product split of marketing spend. Author is testing whether the market share variables are partially explained by other variables and also increases the focus on marketing spend. This model starts with 59 variables.
- Model 4: The final model is intended to be used for practical cost prediction. Author has lagged and excluded some of the variables based on their availability at any given week. This way author can simulate a realistic scenario where the analyst is asked to predict costs of an upcoming campaign and has limited information. All marketing spend share information have been excluded, operational and product opening variables have been lagged by 2 weeks (as at any given week (week 0) for the next week's (week +1) forecasting, the latest full week data is only available from the previous week (week -1) hence the 2 weeks lag) and all market share and economic variables have been lagged by 4 weeks. For this, author reduced the total number of observations from 205 to 201 as the first complete observation is at week 5 instead of week 1. This scenario is also based on stepwise optimisation and includes 86 predictors.

The next step is the **completion of full model analysis** with the goodness-of-fit evaluation. The initial full model contains 99 variables and all predictors are kept in the model regardless of their significance limit. As Figure 27 shows, the selected variables explain 99% of the variance (based on t-test the null hypothesis is rejected, the explained variance is significantly different to 0).



Source: Compiled by author based on SAS output (please find screen-print in Appendix 9)

Figure 27: Linear Regression, 'Full model' evaluation – linearity

Figure 27 also contains information for the **linearity** testing. The visual assessment of the Residuals by Predicted values suggest that although there are some influential data-points, the model satisfies the linearity criteria. Furthermore, the significance level of the slope is also confirmed by the ANOVA table. The same visual analysis was carried out on all predictors (residuals vs. predictors), which assessment highlighted the following three variable groups and required actions. (The applied SAS code and detailed outputs are available in Appendix 9.)

1. Variables with outliers: three variables required the review and correction of the raw data (*OPERATIONS_TD_APPOINTMENTS, OPERATIONS_INT_VISITS, PRODUCTS_CA_UPGRADES_BY_BRANCH*). It was established that one of the data-points was incorrectly extracted.
2. Variable with non-normal distribution: 11 variables (*PRODUCTS_GI_INTERNET_NW_BY_ALL, PRODUCTS_GI_TD_NEW_BY_ALL, PRODUCTS_INV_BR_NW_BY_ALL, PRODUCTS_SAV_BRANCH_NW_BY_ALL, PRODUCTS_SAV_INT_NW_BY_ALL, PRODUCTS_SAVFT_BR_NW_BY_ALL, PRODUCTS_SAVFT_INT_NW_BY_ALL, PRODUCTS_SAVFT_TD_NW_BY_ALL, MARKET_SHARE_SAV_FT_SH_BR, MARKET_SHARE_SAV_FT_SHARE_INT, MARKET_SHARE_GI _BRANCH*) plus all Spend variables required logarithmic transformation.
3. Variables with normal distribution: no action was required in case of this group.

Following the corrections, all variables meet the linearity criteria and author continue with the assessment of the basic assumptions. Author is applying the Durbin-Watson test (DW) to determine whether the error terms are **independent**. Given that the model is based on time-series data a small degree of positive autocorrelation is expected however the large number of predictors can reduce the degree of freedom and the accuracy of prediction, which theoretically can reduce the impact of autocorrelation. As Table 21 shows, the D value is 2.071, which suggest that the error terms are independent. Although the full model is inefficient in terms of number of predictors, it meets the independence criteria.

Table 21: Durbin-Watson Test (Full Model)

Test	Value
Durbin-Watson D	2.071
Number of Observations	205
1st Order Autocorrelation	-0.036

Source: Compiled by author based on SAS output (please find screen-print in Appendix 9)

In order to ensure that the p-values for the t-tests are valid, author is testing that residuals are **normally distributed**. The null hypothesis of the Shapiro-Wilk test is that the data are normally distributed. If the chosen alpha level is 0.05 and the p-value is less than 0.05 then the null hypothesis - which says that the data are normally distributed - is rejected. If the p-value is greater than 0.05, then the null hypothesis is not rejected. The 0.311 p-value of the Shapiro-Wilk test in Table 22 indicates the normal distribution of the error terms.

Table 22: Tests for Normality (Full Model)

Test	Statistic		p Value	
Shapiro-Wilk	W	0.991876	Pr < W	0.3112
Kolmogorov-Smirnov	D	0.051221	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.055079	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.330211	Pr> A-Sq	>0.2500

Source: Compiled by author based on SAS output (please find screen-print in Appendix 9)

The final basic assumption of linear regression is the **homogeneity of variances** of the residuals. In well-fitted models there is no pattern to the residuals plotted against the predicted values therefore heteroscedasticity can be ruled out. The analyst can choose from a series of robust statistical tests, author in current study has chosen the two sample F-test procedure based on the equal split of the residuals by the predicted value where the cut of point is the median value of all predictions.[122]

[122] Group 1 is where the relevant predicted value of the residual is greater than or equal to the median, Group 2 is where the predicted value is less than the median value of all predictions.

In the formal F test, author conducted, the p-value for the F statistic from SAS is compared against the two-sided 0.05 significance level. As Appendix 9 shows, the p-value for the F test using SAS is not significant at α=0.05 (p = 0.6622) so we fail to reject null hypothesis (H_0: $\sigma_1^2 = \sigma_2^2$), with other words, the variances are equal and the model satisfies the homogeneity of variances criteria.

On top of the four basic assumptions author has also assessed the presence and potential impact of **multicollinearity**. The variance inflation factors of the included variables indicated medium multicollinearity across several metrics however in case of only 5 predictors, listed in Table 23 below, the VIF value exceeded 10. Although these variables have not shown correlation at the pre-modelling assessment stage when included to the regression model, they show high correlation as they explain the same phenomena (e.g. most of the media spend was made for current account and was spend on TV advertisement therefore the same factor is explained in two different variables).

Table 23: Test for multicollinearity (Full Model)

Variable	DF	Param. Est.	Std. Error	T value	Pr > |t|	Std. Est.	VIF	Tol.
SPEND_PRODUCT_BRAND_BANK	1	0.0049	1.23	1.23	0.29	1.24	37.68	0.026
SPEND_PRODUCT_CA_BANK	1	0.0049	1.29	1.29	0.29	1.71	70.72	0.014
SPEND_MEDIA_DIGITAL_BANK	1	-0.0048	-1.16	-1.16	0.31	-0.69	12.22	0.082
SPEND_MEDIA_OUTDOOR_BANK	1	-0.0048	-1.16	-1.16	0.30	-1.36	46.52	0.021
SPEND_MEDIA_TV_BANK	1	-0.0050	-1.30	-1.30	0.28	-1.43	47.93	0.021

Source: Compiled by author based on SAS output (please find screen-print in Appendix 9)

It could result inflated standardised estimates and potentially incorrect predictions. Given that the full model only provides a basis for comparison and it will not be used for forecasting or interpretation, author is keeping all variables regardless of the significance of the variable or the multicollinearity. Based on model diagnostics, author concludes that the Full Model has satisfied the basic assumptions of the linear regression. Despite the moderate to high multicollinearity, model is valid although the large number of predictors reduces the accuracy, makes the interpretation difficult and the model inefficient.

The analysis of the baseline or full model is followed by the **completion of the challenger models**. The first challenger model to complete is using all variables from the datamart and applying the stepwise selection method. The variable with the strongest partial correlation (to the dependant variable) with a t value significance <0.05 limit will be selected to enter the model. The exit criteria is the same t<0.05 significance limit.[123] The initial stepwise model required 43 steps to reach a level where all variables left in the model are significant at 0.05 level and there are no other predictors to meet the same 0.05 significance level for entry into the model (the detailed list of variables and model diagnostics are available in Appendix 10). Similarly to the Full Model, Model 1 also requires testing of whether the model satisfies the basic assumptions of regression modelling. The initial model includes 25 variables, with an R^2 of 98.33% and a p-value (<0.001), which confirms that the variance explained by the model is significantly different from 0. Although the linearity criteria are satisfied by the model, as Table 24 shows the Durbin-Watson test indicates a positive autocorrelation of the error terms. Before moving to the next test of the assumptions, the model requires correction.

[123] For more information on the method please see the Selecting Model Optimisation method section within Chapter 3.3.5 Stage V. – Multivariate Linear regression modelling.

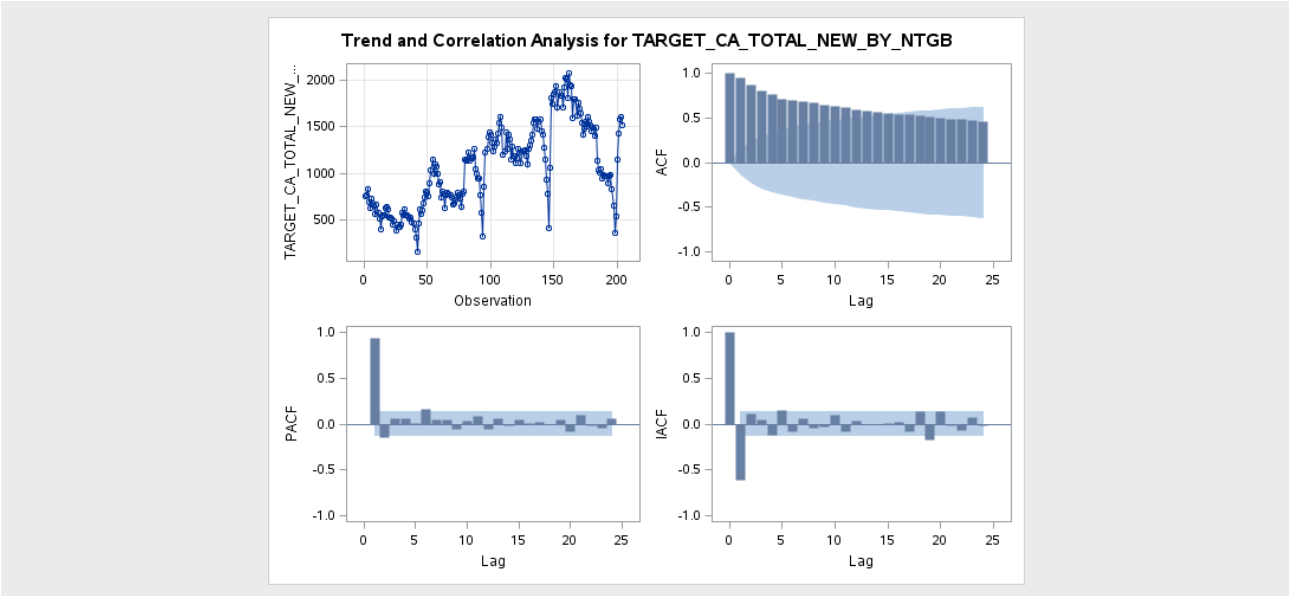
Although the Durbin-Watson statistic is between 1.2 and 1.6 (1.496), which indicates a minor serial correlation, author should understand the required lag before deciding on the approach.

Table 24: Durbin-Watson Test (Initial Model 1)

Test	Value
Durbin-Watson D	1.497
Number of Observations	205
1st Order Autocorrelation	0.248

Source: Compiled by author based on SAS output (please find screen-print in Appendix 10)

For this the Partial Autocorrelation Function (PACF) is applied in SAS which helps to determine the required lag. If there is a significant correlation at the seasonal period (e.g. at lag 4 for quarterly data or lag 12 for monthly data) then the seasonality requires a formal treatment in the model. Seasonality can be handled in a regression model in one of the following ways: a) seasonally adjust the variables; b) use seasonal lags and/or seasonally differenced variables, c) add seasonal dummy variables to the model. In minor cases the simplest way of fine-tuning the model can be achieved by adding lags of the dependent variable and/or lags of some of the independent variables. As Figure 28 highlights (PACF chart) there is a first order autocorrelation present in the model.



Source: Compiled by author based on SAS output (please find screen-print in Appendix 10)
Figure 28: Partial autocorrelation function – Analysis of required lag (Model 1)

Author has manually created the lagged (lag 1) dependent variable and added it to the predictors. The second stepwise iteration of 'Model 1' required 40 steps to reach optimum. The second iteration includes 26 variables with an R^2 of 98.65% and a p-value of <0.001 (Appendix 10). The linearity has been confirmed through the analysis of residuals versus the fitted values (predictions). The analysis of the residuals versus the different predictors did not highlight the need for logarithmic transformation. As Table 25 shows the independence of the error terms is now confirmed, as the new Durbin-Watson statistic (1.699) is above the lower limit (1.6) set by current research. The Shapiro-Wilk test established the presence of normal distribution of the error terms and finally the homoscedasticity is also confirmed by the two sample F-test (the 0.8566 p-value of the test indicates that we fail to reject the null hypothesis and the two test samples of the data has equal variances).

Table 25: Durbin-Watson Test (First Iteration of Model 1)

Test	Value
Durbin-Watson D	1.699
Number of Observations	204
1st Order Autocorrelation	0.145

Source: Compiled by author based on SAS output (please find screen-print in Appendix 10)

However, when assessing the presence of multicollinearity analyst found no evidence for high correlation as the overall VIF figure is 3,754, but there are 3 variables that exceeded the VIF value of 10: the lagged target variable (15.0); the UPL total openings (14,2) and the Credit Card openings (14.4). The most common element of these loan products is the fact that most banks either require their customers to complete the regular repayments to a current account from the same bank or if it is not a requirement then the banks tend to offer discounted rates or other benefits when a credit card or the unsecured personal loan is opened in conjunction with a current account. It highlights that these variables are more likely to follow the same pattern through the cross-sale effect. Despite of the multicollinearity, the effect of these variables is important and author has chosen to keep them in the model.

The second iteration of the full model meets all criteria however for the sample size of 204 observations author aims to reduce the number of explanatory variables further, closer to the 30 observations per predictors, 7-8 variables without the significant loss of meaning. In order to achieve this, three additional challenger models (outlined earlier) have been completed applying the same principles, testing and correction methods. Following the building and completion of the models author summarises the diagnostics and techniques applied for the three other models in Table 26.

Table 26: Diagnostics and techniques applied for model selection

Model Information	Full Model	Model 1	Model 2	Model 3	Model 4
Dependent Variable	Volume of Current Account New to Group Customers				
Modelling technique	OLS				ULS
Variable selection method	Enter all	Stepwise (significance level to enter: 0.5; to leave the model: 0.5)			
Time Periods	Weekly				
No. of input variables	99	99	83	59	86
No. of output variables	99	26	24	18	13
Iterations	1	2 (independence)	2 (independence)	2 (independence)	3 (mixed reasons)
Adjusted R^2 *(F-test significance)*	0.9894 *<0.0001*	0.9865 *<0.0001*	0.9839 *<0.0001*	0.9742 *<0.0001*	0.9284 *<0.0001*
Adjusted R^2	0.9794	0.9845	0.9817	0.9717	0.9234
S.E. of regression *(95% conf. interval)*	3.95%	4.85%	5.56%	6.67%	8.8%
Linearity	Yes (46 log transformations)	Yes (no transformation)	Yes (no transformation)	Yes (no transformation)	Yes (no transformation)
Independence	Yes, DW=2.071	Yes, DW=1.699	Yes, DW=1.611	Yes, DW=1.557	Yes, DW=1.6767
Normality	Yes, Shap.-Wilk p value = 0.3112	Yes, Shap.-Wilk p value = 0.0894	Yes, Shap.-Wilk p value = 0.1426	Yes, Shap.-Wilk p value = 0.0512	Yes, Shap.-Wilk p value = 0.1285
Constant error variances	Yes, F-test p value = 0.6622	Yes, F-test p value = 0.8566	Yes, F-test p value = 0.7189	Yes, F-test p value = 0.0709	Yes, F-test p value = 0.5696
Multicollinearity	Yes, High	No, VIF = 3.8	No, VIF = 3.4	No, VIF = 3.3	No, VIF = 2.4
Autoregression	No, 1st Order Autocorr.= -.036	Corrected, 1st Order AC= 0.145	Corrected, 1st Order AC= 0.167	Corrected, 1st Order AC= 0.183	Corrected with Cochrane-Orcutt
Details	Appendix 9	Appendix 10	Appendix 11		

Source: Compiled by author

The next step is the **selection of the 'champion' model** using the adjusted R^2 and theoretical consideration. When considering the final model, author is selecting from model 1-3 for the

explanation and model 4 for forecasting. The three explanatory models have similar adjusted R^2 values (in fact all three values are within a 1.28% range), which increases the importance of the logical consideration and the analysis of other model diagnostics. Although model 3 contains only 18 variables, which is ideal for interpretation, both tests of the normality and the constant error of variances indicate less significance for the respective test results. Furthermore, the Durbin-Watson test shows an existing weak autocorrelation following the attempt to correct the serial correlation of the error terms. Based on the above highlighted facts, author is excluding model 3 from list of potential champion models. Both Model 1 and 2 indicate model stability. Although the independence of error terms indicates that model 2 is closer to the lower limit compared to model 1, it is important to note that the intention with the champion model is to analyse and understand all key factors accountable for the variance in the new to bank current account customer recruitment. As serial correlation primarily impacts the forecasting capabilities, author is focusing on the multicollinearity, which can lead to false conclusion by inflating the standard estimates. Model 2 has an overall VIF value of 3.4 with no variables above the upper limit VIF value of 10 compared to the 3.8 value in model 1 with 3 highly correlating variables. Taken all model diagnostics into account, author keeps model 2 for the final step. Author has chosen 2 final models; the key difference between the chosen Model 2 and Model 4 is the purpose of the modelling exercise as outlined below:

- **Model 2 is an explanatory (theoretical) model** that incorporates all explanatory variables. There are no lagged input estimators (apart from the target variable) as the ultimate aim is to understand the composition of the variance of the target variable based on a wide range of actual data that became available only during or after the observation has taken place. This method provides an accurate insight and it is useful if the analyst or financial controller is required to provide a post campaign evaluation when the actual market information is available.
- In contrast **Model 4 is a practical forecast model**. When the financial professional is required to conduct a spend forecast or planning exercise, only limited information is available at the point of analysis. The latest data for most of the predictors are either from the previous week or previous month compared to the point of analysis, which normally targets future (next week for example) campaign cost. Model 4 excludes variables that are not possible to obtain at the point of the forecast exercise and has a range of lagged input variables 2 or 4 weeks behind the target period as detailed in Challenger Model 4 from chapter 4.2.5.

Based on the above logical consideration, author has chosen to interpret Model 2, which supports the testing of hypotheses H3.1, H3.2 and H4.1 and used Model 4 for testing hypothesis H5.1. Before the interpretation of the final exploratory model, author is reviewing the regression equation, specifically the parameter estimates, using Table 27 based on the SAS output from Appendix 11. Hypothesis 3.1 states that in order to provide a statistic based and potentially more accurate and efficient campaign cost planning and evaluation method, a mathematical model between the campaign goals, the marketing costs and other contributing factors must exist. To prove this hypothesis we need to find evidence for the existence of a mathematical function (model) between key bank marketing goals[124] and campaign costs with other factors. With other words, the question is whether the regression model contains at least one explanatory variable significantly (linearly) related to the dependant variable. The null hypothesis (H_0) is that: $\beta_1 = \beta_2 = \beta_{max} = 0$ where β is the linear parameter estimate of the predictor. The alternative hypothesis (H_A) is at least one $\beta_j \neq 0$ (for j = 1, 2, max). By reviewing Table

[124] Such as the number of new customers acquired or the level of market share reached per campaign.

27 we can find 24 variables with parameter estimates (β) significantly (where $p < 0.05$) different to 0. 7 of the 24 explanatory variables are media spend related, which proves that a mathematical model between campaign goals, media spend and other factors exist.

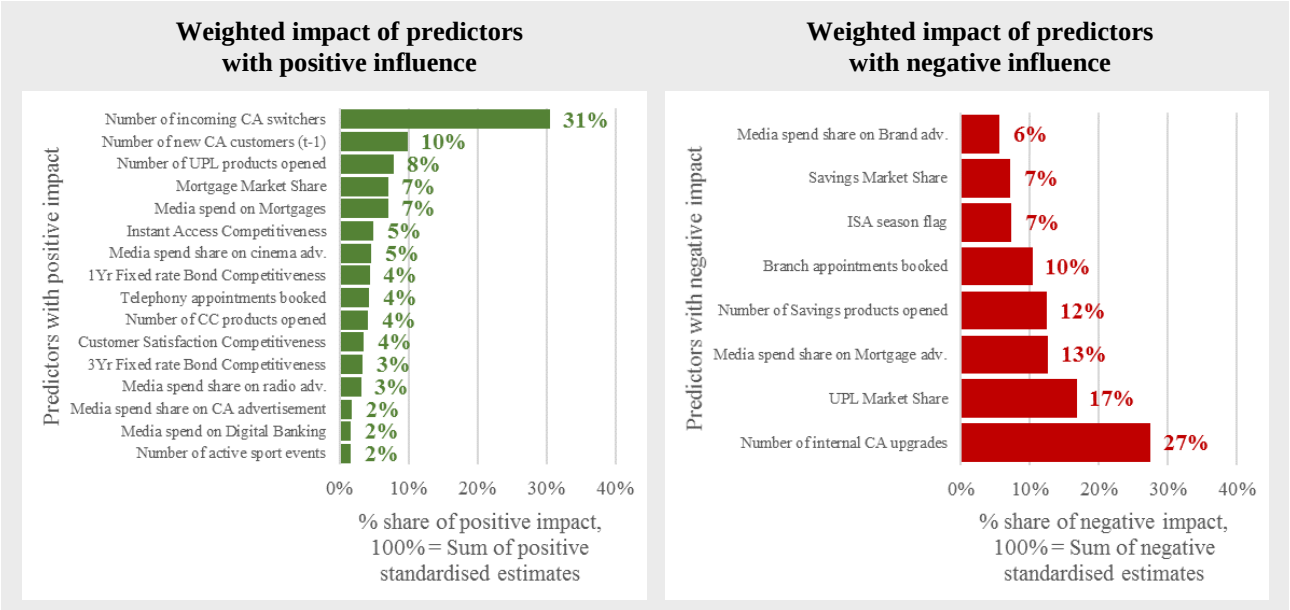
Table 27: The computed Parameter and Standardised Estimates (Model 2)

Model Variables	Parameter Estimates	T value	p value	Standardised Estimates
Intercept	243.07459	1.05	0.02949	0
TARGET_CA_TOTAL_BY_NTG_LAG1	0.18688	5.84	<0.0001	0.18686
OPERATIONS_BR_COMPL_APPOINTMENTS	-0.00998	-5.87	<0.0001	-0.08828
OPERATIONS_TD_APPOINTMENTS	0.04096	3.64	0.0004	0.08081
OPERATIONS_CUST_SAT_COMPETITIVE	37.33324	3.83	0.0002	0.06765
PRODUCTS_CA_UPGRADES_BY_ALL	-0.05498	-7.82	<0.001	-0.23333
PRODUCTS_CC_TOTAL_NEW_BY_ALL	0.02763	3.42	0.0008	0.07923
PRODUCTS_SAV_TOTAL_NEW_BY_ALL	-0.01436	-4.42	<0.0001	-0.10589
PRODUCTS_UPL_TOTAL_NEW_BY_ALL	0.18467	4.47	<0.0001	0.14826
MARKET_SHARE_SAV_OTH_TOTAL	-1064.57559	-2.63	0.0092	-0.06136
MARKET_SHARE_UPL_OTH_TOTAL	-12735	-6.29	<0.0001	-0.1426
MARKET_SHARE_MR_OTH_TOTAL	11718	5.05	<0.0001	0.13344
MARKET_COMP_SAV_ISA_1_YR_FR	60.09034	3.4	0.0008	0.08308
MARKET_COMP_SAV_BOND_3_YR	50.67525	3.57	0.0005	0.0635
MARKET_COMP_SAV_INSTANT_INTERNET	115.07185	6.56	<0.0001	0.09345
MARKET_COMP_CA_SWITCHERS	0.46494	15.31	<0.0001	0.57935
ISA_SEASON_FLAG	-150.49415	-3.43	0.0008	-0.0628
SPORTS_EVENTS	66.21749	2.86	0.0048	0.03095
SPEND_PRODUCT_BRAND_SHARE	-445.60663	-4.22	<0.0001	-0.0475
SPEND_PRODUCT_CA_SHARE	151.30542	2.6	0.0101	0.03497
SPEND_PRODUCT_MORTG_BANK	0.00109	5.99	<0.0001	0.1331
SPEND_PRODUCT_MORTG_SHARE	-774.24526	-4.86	<0.0001	-0.10727
SPEND_MEDIA_CINEMA_SHARE	252.92052	5.88	<0.0001	0.08688
SPEND_MEDIA_DIGITAL_BANK	0.00021797	2.7	0.0075	0.03113
SPEND_MEDIA_RADIO_SHARE	564.65089	5.56	<0.0001	0.05982

Source: Compiled by author

It also provides evidence for Hypothesis H3.2, which states if such mathematical model exists then the start-to-end process to support analysts and controllers to conduct marketing campaign cost prediction and evaluation analysis should be captured. The statistical process, from the conceptualisation of the modelling aim to the statistically validated regression equation, can be captured on simple process diagrams to support financial professionals. As the valid regression model for H3.1 was built using the Process Diagram of Stage I.-VII., it partially confirms that the process can be successfully captured and used for marketing campaign evaluation. For full confirmation author is required to prove that the practical model (Model 4) can also be used for effective cost forecasting. The final hypothesis that can be confirmed by the analysis of Table 27 is H4.1 which states that there is a linear relationship between the campaign goals, the campaign spend and most of the other contributing factors (hence multivariate linear regression is an adequate method). The question therefore is whether any of the media spend variables are significantly (linearly) related to the dependant variable? It presents the following hypothesis structure: H_0: We can only establish a model where $\beta_1 = 0$ (the null hypothesis means that none of the media spend variables are significantly, linearly related to the campaign target) and H_A: We can establish a model where $\beta 1 \neq 0$ (the alternative hypothesis means that at least one media spend variable is significantly related). As previously stated, 7 of the 24 predictors are campaign spend related with a parameter estimate (β) significantly greater than 0 therefore we fail to reject the null hypothesis and conclude that the target variables, the campaign cost and other factors can be assessed with a linear model.

The final step is the **interpretation of the model**. The selected champion model explains 98.38% of the variance of the target variable. It indicates a good explanatory power however the R^2 statistics can be inflated by the large[125] number of variables. Using the standardised parameter estimates from Table 27 we can rank all 24 variables based on their influence on the model. An influence is considered positive if the standardised estimate is a positive value, which in this case is considered as a factor that result increment in the total number of new to Bank A customer acquired during the campaign period. Following the same logic, an influence is considered negative if, the standardised estimate is negative and the factor results reduction in the total number of new to Bank A customers, with other words, these are treated as factors that drive new potential business out. Based on the analysis of Figure 29 the absolute competitiveness of the current account product proposition variable[126], measured by the number of current accounts weekly switched into Bank A through the dedicated switcher service in the United Kingdom, is the most influential positive driver. It is providing 31% of the weekly New to Bank A incoming volumes. It is followed by the impact of the sales performance from the previous week with 10% of all increment.



Source: Compiled by author based on SAS output (please find screen-print in Appendix 11)

Figure 29: Weight impact of predictors split the direction of the influence

The most influential negative driver is the current account upgrade. This driver requires further analysis to establish the link between the predictor and the target variable. Author's underlying assumption is that this predictor is connected to the target variable through seasonality. When the product offering favours to existing customers (for example in ISA campaign season the existing customers with the flagship current account receive additional benefits), it is more likely that existing customers upgrade rather than new customers purchase. To test this assumption author has looked for other negative influence drivers and the correlation matrix to establish if the predictors are explaining the same phenomena. The ISA flag and the number of Savings products sold variables confirm the hypothesis as all three variables have negative impact on the new customer recruitment figures and have a weak correlation. Generally, savings seasons such as ISA or the Bond seasons are targeting savers. Savings customers are likely to be at a later life stage where they can afford to separate a part

[125] Author was aiming to reduce the number of variables to 30 observations per predictor (7-8 variables). Anything more than this is considered as 'large number of variables'.
[126] MARKET_COMP_CA_SWITCHER variable.

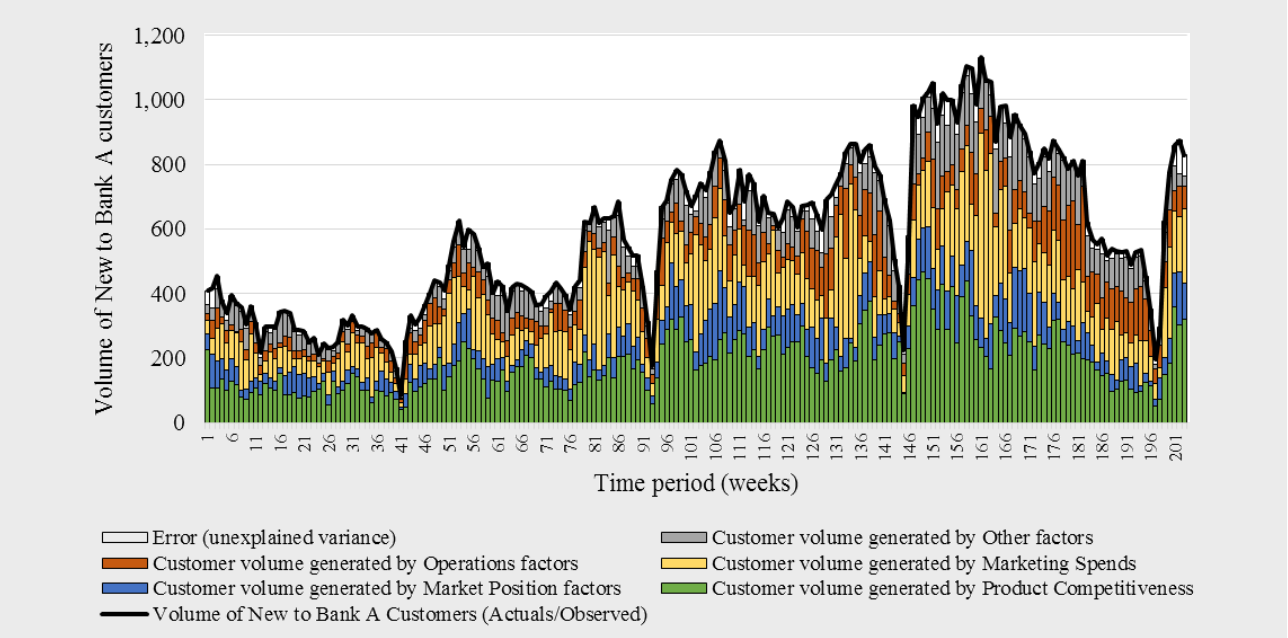
of the regular income for savings, it is especially true with Bond customers who are typically well established customers with set financial preferences. These customers have less likely have a financial need for Bank A's current account, especially if they are existing customers, but they have for a savings instead. Therefore products, targeting this customer group are on offer, it is less likely to increase the number of new to group customers but to potentially increase product upgrades for beneficial rates. This theory is also supported by the negative impact of the Booked Branch Appointment as typically Branch is the preferred channel of the elderly savings customers over the internet, which drives most of the current account switcher business. Further strategic observation an analyst can make is that unsecured personal loans and credit cards are positive enablers of the new to Bank A current account stock increment. It would require further analysis to determine whether the current account results the credit card cross-sale or it is the other way around but with unsecured personal loans it is common banking practice of the financial institutions to request the monthly repayments to be paid from a current account managed by the same financial institutions.

Following the analysis of the weighted standard estimates author is analysing the influence of the different variables at given observations. With the use of the differential standardised estimates or DFBETAS[127] function, it is feasible to quantify how much change an observation would make on a predictor, with other words, what number of new to Bank A customers each predictor resulted at a given period of time. In order to manage the analysis more efficiently author is grouping the 24 output variables into 5 main variable groups:

- **Operations:** focused on variables from the model that are related to the operational activity: *OPERATIONS_BR_COMPL_APPOINTMENTS, OPERATIONS_TD_APPOINTMENTS, OPERATIONS_CUST_SAT_COMPETITVE*
- **Product Competitiveness:** centred around variables measuring the competitiveness of the product proposition. These are measures either relative to a market position or an absolute volume of business moved out from competitors: *PRODUCTS_CA_UPGRADES, PRODUCTS_CC_TOTAL_NEW, PRODUCTS_SAV_TOTAL_NEW, PRODUCTS_UPL_TOTAL_NEW, MARKET_COMP_SAV_ISA_1Yr, MARKET_COMP_SAV_INST_INT, MARKET_COMP_CA_SWITCH, MARKET_COMP_SAV_BOND_3_Yr,*
- **Market Position:** it contains all market share related variables: *MARKET_SHARE_SAV_TOTAL, MARKET_SHARE_UPL_TOTAL, MARKET_SHARE_MR_TOTAL*
- **Marketing Activity:** in current model the marketing activity variables are related to the absolute and relative to marketimg media spend: *SPEND_PRODUCT_BRAND_SHARE, SPEND_PROD_CA_SHARE, SPEND_PROD_MTG_BANK, SPEND_PROD_MTG_SHARE, SPEND_MEDIA_CINE_SHARE, SPEND_MEDIA_DIGI_BANK, SPEND_MEDIA_RADI_SHARE,*
- **Other:** the last category includes the error term and the impact of different flags: *ISA_SEASON_FLAG, SPORT_EVENTS*

Figure 30 shows the grouped contribution of the predictors towards the total number of new to Bank A CA customers by their different weekly impact.

[127] "*The measure that measures how much impact each observation has on a particular predictor is DFBETA. The DFBETA for a predictor and for a particular observation is the difference between the regression coefficient calculated for all of the data and the regression coefficient calculated with the observation deleted, scaled by the standard error calculated with the observation deleted*." Definition was accessed on 30/10/2016 from: http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg_self_assessment_answers2.htm



Source: Compiled by author based on SAS output

Figure 30: The impact of predictors on each observation (DFBETAS)

The graphical analysis proves that the Product Competitiveness of the product mix is the key contributor group. When the strategic activity is overlaid the graph, it confirms that Bank A in fact had a strong savings and current account proposition between time-period 46 and 56; a strong credit card and current account offering between 96 and 104 and a strong mortgage and current account packaged deal between period 146 and 151 as the factors related to the product proposition are resulted more accounts than the other groups. The Marketing Spend has the second highest influence. The assessment of the differential standardised estimates show greater contribution from the marketing spend factors in current account campaign periods such as student current account or mortgage seasons than at other times. These are the times when most of the current account related marketing activity is held and paid for. The factors related to the market position has the least impact over time. The largest contribution was registered around the time of the mortgage and current account packaged deal, which can be an indirect impact of the increased brand awareness of the potential mortgage customers on the market or the promotion of the intermediary channel.

As the above analysis demonstrates, the use of the Standardised and Differential Standardised Estimates enables analysts and financial professionals to determine the key factors (and their weights) of the marketing campaign success from a management control perspective. This partially confirms hypothesis 4.2, which states that using the modelling database and the multivariate linear model, the key factors that determine success of a bank marketing campaign (such as number of new customers and size of market share per campaign) can identified with their weight (or contribution) to the success. H4.2 is partially proven and further evidence will be provided in Chapter 4.4.

4.2.6 Stage VI. – Cross-Validation the model

Following the completion and the initial diagnostics of the model assumptions, author is cross-validating the model to test whether it is appropriate to generalise the interpretation and to ensure that

the prediction is accurate for the cost modelling exercise. At current point of time an additional 26 weeks data[128] are available. Author has chosen to complete two sets of cross-validation tests:

1) Predict values based on new dataset and compare the prediction against actuals;
2) Refit the model on a validation partition[129].

1) Comparison of prediction against actuals: For the final predictions author is using the structural predictions[130] and the prediction on error terms[131] to correct the serial correlation. The predicted values are then compared against the actuals using a series of indicators to understand the prediction accuracy. Author has calculated the Mean Absolute Deviation (MAD)/Mean ratio, the Mean Absolute Percentage Error (MAPE) metric and the Pearson Correlation for testing. MAD measures the size of the error in the prediction units. It is calculated by the $\sum_n^1 |y - \hat{y}|$ formula where n is the number of observations, y is the actual value and $\hat{y}$ is the forecasted value. This is then divided by the mean of the actuals. A ratio close to 0[132] is desired as it would provide evidence for an accurate forecast. MAPE measures the size of the error in percentage terms. The following formula is applied: $\left(\sum_n^1 \frac{|y-\hat{y}|}{|y|}\right) * 100$. Although this metric is scale sensitive, the volume of customers is not low-volume[133] data therefore the application of the test is appropriate. Table 28 summarises the cross validation diagnostic metrics.

Table 28: Cross-Validation Diagnostics (predictions vs. actuals)

Measure	Value
Number of predicted periods (weeks)	24
Observed Mean	802.773
Predicted Mean	744.92
Mean Absolute Deviation (MAD)	58.68
MAD/Mean Ratio	7.3%
Mean Absolute Percentage Error (MAPE)	7.1%
Pearson Correlation	0.748

Source: Compiled by the author

Both the MAD/Mean ratio and the MAPE test results are below 10%, which highlights the imperfection of the forecast modelling. However, it proves that the model still can be used for prediction. It is important to highlight that when only modelled estimates are used instead of forecasts (residuals are calculated based on observations and modelled values instead of minimum variance predictors) then the model accuracy significantly increases (both the MAD/mean ration and the MAPE reduce to 2.7% with a 87% correlation), which means the model in fact can be used for campaign cost evaluation process.

2) Refit the model on a validation partition: in case of cross sectional data, analyst can simply split the data into a training and validation partition by random selection. When time series data is used, the analyst should use the forward chaining or canonical cross validation method when creating the

[128] Weekly data between 08-Feb-2016 and 08-Aug-2016
[129] For detailed methodology please see Chapter 3.3.6 Stage VI. – Cross-Validation of the model.
[130] Structural prediction: prediction based on the estimated intercept and the estimated parameter on the independent variable.
[131] Prediction on error terms: the error predictions are calculated based on the lagged residuals (1 period lag in the current model) of the structural prediction and the actuals. Where the residual is missing it is replaced by the minimum variance predictor. For more details please see next (Cost prediction and intervals) chapter.
[132] For example: MAD/ratio below 10%.
[133] Low-volume data in this context means that the actuals and forecasts have values between 0 and 1.

partitions due to the serial correlation in the data. Author has chosen a 3-fold cross validation method where the training model of each fold contains all observations from the respective number of consecutive year(s) then the models are validated on all observations from the next available year. This way the training dataset will not contain information that occurs after the test set. Once the folds have been created, the average mean squared errors (MSE) of the folds and the overall MSE is analysed to determine prediction stability of the chosen variables. Table 29 summarises the approach and the test results.

Table 29: Cross-Validation Diagnostics (3-fold validation)

Test	Training Dataset	Validation Dataset	MSE	%
Fold 1[134]	Year 1 (50 records)	Year 2 (50 records)	50,930	127%
Fold 2[135]	Year 1, 2 (100 records)	Year 3 (50 records)	43,730	109%
Fold 3[136]	Year 1,2,3 (150 records)	Year 4 (50 records)	44,334	111%
		Based on the average error:	46,331	116%
		Full model:	39,990	100%

Source: Compiled by the author

Although Fold 1 has 27% higher MSE compared to the full model, it is within 11% of the average test result, which suggest the overall model is stable. Furthermore, the full model has the lowest mean standard error which also suggested it is the most accurate prediction model based on the available modelling dataset. Author therefore has concluded that the model could be used for cost prediction and the results were adequate for generalisation.

4.2.7 Stage VII. – Cost prediction and intervals

Following the final model build and cross-validation, the analyst can start to apply the model for cost modelling exercise. Depending on the time and data availability the analyst can use the model for two main purposes (from the cost modelling perspective):

1. Prediction: using new data and the model the analyst can predict the expected outcome (e.g. number of customers recruited within the campaign period) for certain time periods or campaigns. Then this information can be used for post campaign cost evaluation to understand whether Bank A over or underspent during the campaign. Furthermore, the analyst can determine the weight of the factors contributed to the result, which helps the next planning and control cycle.
2. Cost forecasting: using forecast data for the variables of the regression equation and campaign target figures for the target variable, the analyst can determine the required media spend with some caveats. Depending on the model, the analyst can determine the spend requirements by media type.

Author will demonstrate both use cases. As mentioned in the previous chapter, at current point of time an additional 26 weeks data[137] are available, which will be used for the prediction to support the analysis (or 'check' function of the PDCA cycle) of actuals and targets. Furthermore, author is using the weekly forecasts and targets of the same variables instead of the actuals similarly to the data was available for traditional planning method. It helps to replicate a true cost forecasting exercise.

[134] Training: 50 consecutive observation records between 2-Apr-2012 and 11-Mar-2013, Validation: 50 consecutive observation records between 18-Mar-2013 and 24-Feb-2014
[135] Training: 100 consecutive observation records between 2-Apr-2012 and 24-Feb-2014, Validation: 50 consecutive observation records between 03-Mar-2014 and 16-Feb-2015
[136] Training: 150 consecutive observation records between 2-Apr-2012 and 03-Mar-2015, Validation: 50 consecutive observation records between 23-Feb-2015 and 01-Feb-2016
[137] Weekly data between 08-Feb-2016 and 08-Aug-2016

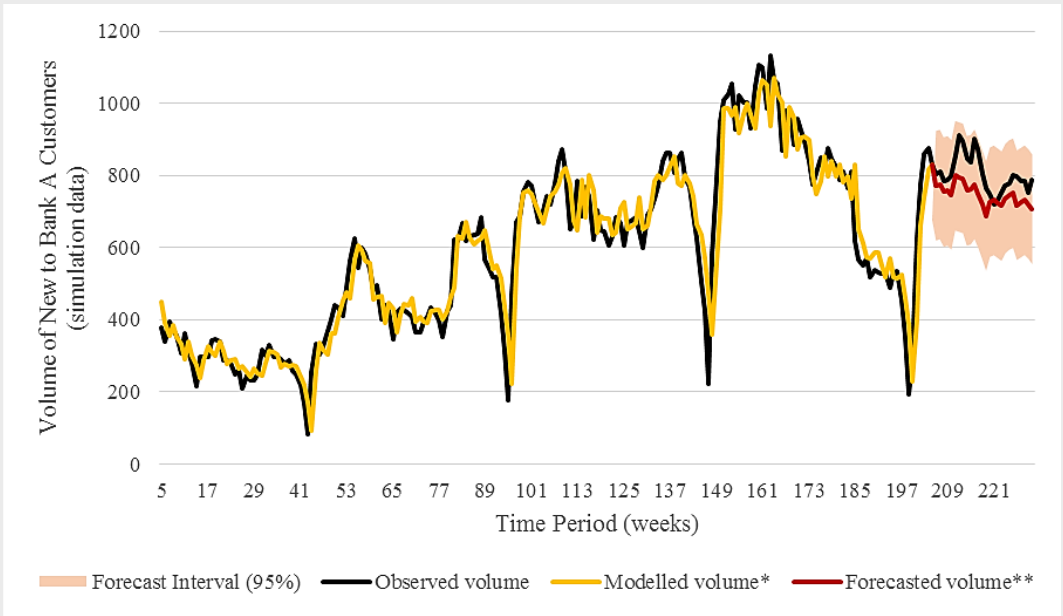
Prediction: the following regression equation can be extracted from the SAS software output:

Table 30: The computed regression equation (Model 4)

Variable (Target and predictors)	Parameter Value
New to Group CA customer volume =	-1436.29521
+ *Total CA media spend (£)* ×	0.00005
+ *Number of active TV campaigns* ×	32.54719
- *Number of switcher CAs (t-2)* ×	0.02688
+ *Customer satisfaction competitive score (t-4)* ×	38.08070
+ *Economic Stability FACTOR (t-4)* ×	453.08608
+ *Brand warmth index (t-4)* ×	1211.29985
+ *Number of Credit Card sold (t-2)* ×	0.04851
+ *Number of Investment sold (t-2)* ×	0.04588
- *Number of Mortgage sold (t-2)* ×	0.09843
+ *Mortgage market share percentage (t-4)* ×	30617.78903
(AR1 – First order Autoregressive Coefficient = -0.79475)	

Source: Compiled by the author based on SAS output (please find screen-print in Appendix 11), t-2 refers to a -2 weeks lag, t-4 to a -4 weeks lag, for more details please review Challenger Model 4 from chapter 4.2.5.

The first step is to calculate the structural predictions based on the estimated intercept and the estimated parameters of the independent variables. In case of time-series data the next step is to adjust this prediction with the prediction on error terms. Error predictions are calculated based on the lagged residuals (1 period lag in the current model) of the structural prediction and the actuals. Where the residual is missing, it is replaced by the minimum variance predictor, which is the lagged residual multiplied by the autoregressive coefficient (AR1). In case of a second (or more) order autoregressive model, the minimum variance predictor is the sum of respective residual lags multiplied by the respective coefficients (AR2, ARx). Figure 31 shows the combined results of the prediction (modelling) and the forecasting exercises.



* Modelled volume: volume prediction where the residual between the observed and structural prediction is known and it forms the basis of autoregressive correction.

** Forecasted volume: volume prediction based on the minimum variance predictor.

Source: Compiled by author

Figure 31: The observed, modelled (predicted) and forecasted volumes over time

When the residuals between the observed and the structural predictions are known[138], the model generates an accurate[139] prediction (modelled volume) after the correction of the serial correlation. As the cross-validation results highlighted, the model is also able to accurately forecast the volume of new customers recruited weekly at given marketing spend and other contributing factor level for 26 new period units. The root mean squared error (observed-forecasted) versus root mean squared value (of observed) ratio is 8.3% over 26 weeks and the correlation is 75% when compared with actuals. The prediction interval is ±151 customers per week, which is ±20.3% range at 95% confidence level. Although the model is accurate, it is important to highlight that the forecast will only be accurate if the estimates or forecasts of the contributing predictors are accurate.

Cost forecasting: The most appropriate way to create a cost prediction model is to apply the regression model building process from start to end as outlined in Chapter 3, aiming the prediction of the cost variable. An alternative option is to use the regression equation obtained from the recently built customer recruitment prediction model alongside with the forecasted predictors and the volume targets as the target variable. This method presents a potential issue. Using the annotations from Chapter 3.3.5 the mathematical challenge can be described in the following way: on completion of the regression analysis, author obtained the following equation:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i + \varepsilon$$

If we assume that y is the targeted customer volume for one period (e.g. 1 week), x_1 is the spend variable we aim to forecast, x_i represents the forecast of all other variables by the chosen period and the Betas (β) are known from the regression equation then our aim is to identify x_1. The problem is that there is no information about the size of the error. If we rearrange the equation in a way that all known variables are on one side and unknowns are at the other side we can see the problem:

$$\beta_1 x_1 + \varepsilon = y - \beta_0 - \beta_i x_i$$

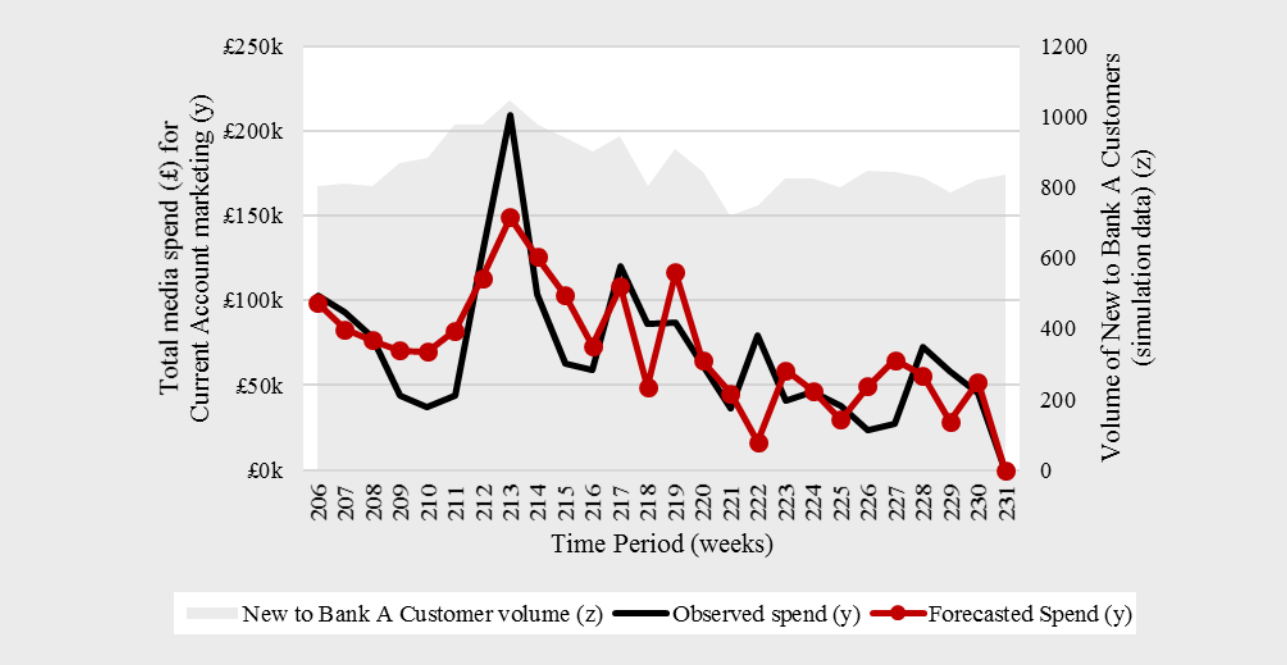
When the regression equation is applied to calculate the required marketing spend, the analyst should find a way to separate the error terms from the impact of spend. There are multiple options however all of them present an issue around accuracy and therefore the analyst should make a decision whether to create a new regression analysis and model the cost variable from start to end or to use the existing model and caveat the results:

- Option 1: If the error terms relative to the model are within the accepted level from thr perspective of the cost model (e.g. the RMSE[140]/RMS[141] ratio is below 1%) then the analyst can choose to estimate the cost including the error terms. Depending on the influence of the cost variable on the variance of the target variable potentially it can overestimate the required spend significantly.
- Option 2: If the analyst would like to forecast the marketing spend requirement of the next period, only then the ratio of the error terms vs. the forecast variable $\frac{\varepsilon_{t-1}}{\beta_1 x_{1_{t-1}} + \varepsilon_{t-1}}$ can be obtained from the previous (t-1) period and used as a proxy to separate the error.
- Option 3: If the analyst would like to forecast multiple periods then either a moving average of a set number of periods or the average of the error terms vs. forecast variable ratio of the overall model can be obtained and used as an error ratio parameter. The issue it presents is the

[138] Between time periods (weeks) 5 and 205.
[139] Based on the 95.4% correlation between the observed and predicted value.
[140] Root-mean-square error
[141] Root-mean-square (of the observed values)

reduction in the influence of the cost predictor, which means that the spend requirements will be over and underestimated in some time periods. However, this overall impact should decrease as the number of time periods to be forecasted is increased.

Author has chosen to use the current regression equation (instead of creating a new model targeting the cost variable) and to use Option 3 (as the RMSE/RMS ratio is 8.3%) for hypothesis testing. The analysis of the error terms vs. the forecast variable have shown that the error term covers 97.46% of the sum on average. When this ratio is applied, the calculation can be completed. For testing purposes the additional 26 weeks data is used as forecasted predictors and the actual new customer recruitment as the target. This way author can compare the forecasted cost values against actuals. On Figure 32 the pale grey area graph shows the number of new to Bank A customers weekly for reference. The black bold line represents the weekly media spend on current accounts by Bank A and the red line is the forecasted media spend using the regression equation and option 3.



Source: Compiled by author

Figure 32: Observed and Forecasted Current Account Media spend

Following the visual inspection of the different media spend graphs, author concludes that the two trends follow similar pattern however further statistical evidence is required to prove or reject Hypothesis 5.1. This hypothesis states that the linear equation extracted with the multivariate linear model can be used for cost forecasting and identifying required cost levels for new customer recruitment targets in the system of other contributing factors for future campaigns. The hypothesis is proven if cost values forecasted by the model for new observations are statistically significant (with other words, the correlation between the forecasted and actual spend values is over 65% and the variances of the forecasted and observed populations are equal at the 95% confidence level). Table 31 details the summary statistics of the correlation between the two population and the results of the test of equal variance. The summary statistics highlight a strong (75.4%) correlation between the forecasted and observed media spend furthermore the F value is lower than the F critical value therefore we fail to reject the null hypothesis of the F-test and can conclude that the two sets of spend values have equal variance at 95% confidence level.

Table 31: Correlation diagnostic and the test of equal variances for H5.1

Measure	Observed Spend	Forecasted Spend
Sum of values	£1,782,236	£1,830,912
Mean	£68,548	£70,420
Variance	£1,777,851,786	£1,238,606,814
Observations	26	26
dF	25	25
F	1.435364125	
P(F<=f) one-tail	0.186151898	
F Critical one-tail	1.955447207	
Correlation	0.753535822	

Source: Compiled and calculated by the author

This confirms Hypothesis 5.1. Practically, it means that a financial analyst or accountant can in fact forecast media spend accurately if the new customer recruitment targets and the forecast of the contribution factors, either by using the current research or building a new regression model internally, are known. This offers another technical approach for cost planning. As the forecast model was built using the Process Diagram of Stage I.-VII., it fully confirms that the model process for campaign evaluation and also for effective cost forecasting can be successfully captured, which provides the final proof for hypothesis H3.1.

4.4 KEY FACTORS THAT DETERMINE THE SUCCESS OF BANK SALES CAMPAIGNS

After the successful completion of one full model, which was aiming to explore the relationship between the new customer recruitment activity of Bank A and the the marketing spend with other key explanatory variables, author has used the same methodology described in Chapter 3 and 4 to complete 9 additional full, validated and optimised exploratory regression models. The additional models were focusing on different target variables however were using the same predictor set as the full model, they were also built by using stepwise regression and only the champion model that satisfied all regression assumptions with the highest adjusted R^2 relative to the number of predictors were selected. The summary of model diagnostics for all final champion models in each additional modelling scenarios are available in Appendix 12.

The following 10 models (including the final model from Chapter 4.3) have been created to determine the key success factors of Banking, Savings and Credit Card sales campaigns:

- Model 1 targeting new to Bank A customers who purchased current account;
- Model 2 targeting new to banking products, existing customers who purchased a current account;
- Model 3 targeting existing customers who purchased a new current account;
- Model 4 targeting all (existing and new) customers who purchased a new current account;
- Model 5 targeting all (existing and new) customers who purchased a new credit card;
- Model 6 targeting all (existing and new) customers who purchased a new savings product;
- Model 7 targeting current account new business (campaign) market share;
- Model 8 targeting current account stock market share;
- Model 9 targeting credit card new business (campaign) market share;
- Model 10 targeting savings new business (campaign) market share.

Table 32 summarises the common and model type specific (product related and market share related) variables. The most **common variable** appearing in 7 out of the 10 model is the current account

competitiveness. The analysis of the standardised estimates is highlighting that the strong current account proposition predictor is a positive driver except in case of the savings model. To understand this negative influence on the savings volumes, the context of current account proposition becomes an important factor. The flagship product of Bank A, similarly to other providers' flagship products, offers higher interest on lower balances (below £25,000) compared to the same balance tier rates for savings, especially instant access accounts. Therefore, savers with lower balances are more likely to take out the current account instead of a savings account hence the inverse influence. Another common variable that mainly impacts the current account and credit card related models is the Credit Card Market Share, which has a higher influence on the product related models.

Table 32: Common and model specific factors

	Weighted impact of factor *(Average across modelling groups)*		
	Product Related Models (6)	Market Share Related Models (4)	
CA Product Competitiveness (Switcher volume)	19%	3%	**Common Factors**
Credit Card Market Share	12%	6%	
Economic Stability FACTOR	4%	6%	
1 Year ISA Product Competitiveness	2%	4%	
Number of Working days	2%	2%	
Instant Access Product Competitiveness	2%	2%	
Credit Card Product Sales	3%	7%	
Number of Telephony Appointments Booked	1%	4%	
3 Years Bond Product Competitiveness	1%	3%	
ISA Season Flag	2%	2%	
Number of active TV campaigns	1%	2%	
Current Account Media Spend	1%	3%	
UPL Market Share	3%		**Model type specific factors**
Mortgage Market Share	3%		
Operations – Internet Visits	2%		
Media Spend - Mortgages	1%		
Media Spend – Digital	1%		
Media Spend Share – Radio Spend	1%		
Media Spend Share – Brand Spend	1%		
Current Account Market Share		15%	
Current Account Openings by new to Bank A custs.		2%	
Media Spend Share – Cinema Spend		2%	
1 Year Savings Bond Competitiveness		2%	

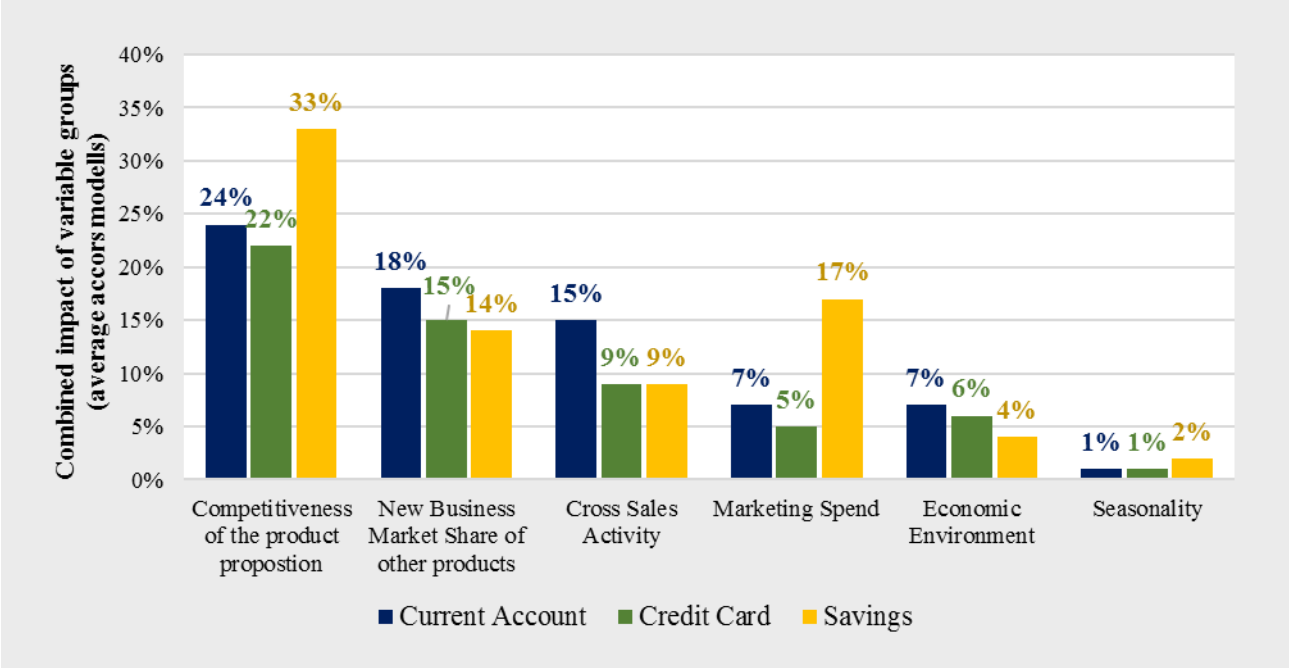
Source: Compiled and calculated by the author

The third most common variable is the Economic Stability FACTOR created at the reduction of dimensions stage. This variable incorporates the gold price, average commercial bond rate and the consumer confidence index predictors that determine whether customers are likely to cut expenses and large purchases or invest and look actively for loans and credit cards. This variable mainly impacts the market share figures but it is also a contributing factor to the current account related models. It shows that the economic area has a quantifiable impact on product sales.

The combined analysis of the models highlighted some model type **specific variables**. The UPL and Mortgage Market Share predictors have a positive impact on all product models (except the Savings model), which highlights the importance of cross-sales activity and packaged deals. Most of the media spend variables are influencing the product oriented models but not primarily the market share models. The Current Account market share has the largest impact on the market share models

however this is skewed by Current Account stock market share. This shows importance of the current account position and the influence on other market shares.

In order to draw a conclusion that can support both the scientific community and the financial controllers, the output of the model is needed to be **generalised**. To achieve this, the output variables have been merged into categories and the impact percentage of each predictor per model calculated by the division of the absolute value of the standardised estimate for a given variable by the sum of all absolute standardised estimates of the model, have been averaged across the models focusing on the same product (for example both product and market share related models current accounts or credit cards or savings). The results have been summarised by the following bar chart on Figure 33.



Source: Compiled by author

Figure 33: The weight of the generalised and grouped campaign factors by product type

Based on the analysis of Figure 33 the competitiveness of the product proposition has the highest influence on the target variable (which can be the number of customers purchasing a modelled product in a campaign period or the level of market share reached at a given campaign). This variables group is centred around the market competitiveness scale variables, which represent the rate ranking for credit cards and savings products and the overall competitiveness of the current account proposition by measuring the absolute number of current account bank switcher customers. The second group is the new business market share of other products. This group shows the importance of a strong product portfolio at Bank A. Strong campaign performance (or high new business market share) of other products such as personal loans and mortgages are highly advertised and in case of competitive rates are also promoted by intermediaries, which ultimately increases the brand and product awareness and can be translated to actual product openings. This group should also be analysed with the cross sales activity. A significant part (15% for current accounts) of the new accounts are opened in junction with another products. Although it would require further analysis to understand the main product in a cross-sale, in some cases (such as unsecured personal loans) there is an increased likelihood for current accounts to be the secondary source. The next group is the marketing spend. Although there is a significant variation on a product group level, the marketing spend group is showing the combined total influence across all spend related metrics. The marketing spend has the highest impact on the

Savings campaigns (17%) and the lowest on Credit Cards (5%). A possible explanation for the difference is in relation to the economic stability factor group. Savings products are less likely to be influenced by economic factors compared to Banking and Credit Cards. This can be a factor of the customer profile and/or a factor of low savings rate across all providers during the modelling period. As discussed in the previous chapter, savers are more likely to be well established people with high enough income to save (enough) money to open a savings account or elderly customers with lifetime savings. If the customer profile is the main driver then the model suggests that these customers are more likely to be influenced through marketing when it comes to savings product purchasing. The other option is the low interest environment. Due to similar propositions across most of the providers, the importance of marketing and therefore the marketing spend from Bank A's perspective is considerably higher compared to the two other product groups. Unfortunately, the model does not provide sufficient information to state that one or the other theory is true but it is still a significant result.

There is an obvious question around seasonality. The seasonality across the 10 models is only accountable for 1.7% of the variance on average with products where the seasonality would normally be expected to play a vital role. Author has built 10 sub models (a model per each target variable) where only the seasonality related variables[142] are included as predictors. The aim is to understand the explained variance by these predictors. The SAS output shows that the ISA season flag is part of all models with the average R^2 value of 14% for current account models; 14% for credit card models 14%; and 38% for the savings model (the saving model includes the ISA and the Bond season flags). Then the question is where this explanatory power has gone from the final models when the additional predictors are included? The simple answer is that the other variables are also partially explaining the seasonality. To check this assumption author is analysing the relationship of these seasonality variables with the rest of the predictors. Although the variance inflation factor did not indicate collinearity that requires attention, when the correlation matrix is examined amongst the variables (Appendix 6) we can see an emerging pattern of weak to medium correlation across banking, savings and credit card openings and market shares as well as with some of the media spend variables. It indicates that the seasonality is already described by other variables and only a small proportion of the explained variance is explained exclusively by the seasonality flags.

An additional, interesting fact has been discovered during the model analysis. The average temperature has been identified as an estimator for savings models (only). This variable has 7% influence on average. The relationship is direct, which suggest the warmer the weather is the more savings products are sold. This may well be a very simplistic, out of context interpretation and when the variable is closely examined and put into contexts a more likely explanation has emerged. The average temperature variable shows a weak to medium correlation with the bond and ISA season flags which suggest that the average temperature is partially explaining the same seasonality factor. Savers are more likely to save in spring and autumn time and spend some of the savings (or have a need to open a savings product) during summer and the winter festive season. Although other factors can also influence the fact that it has no impact on other products, it supports the statement above.

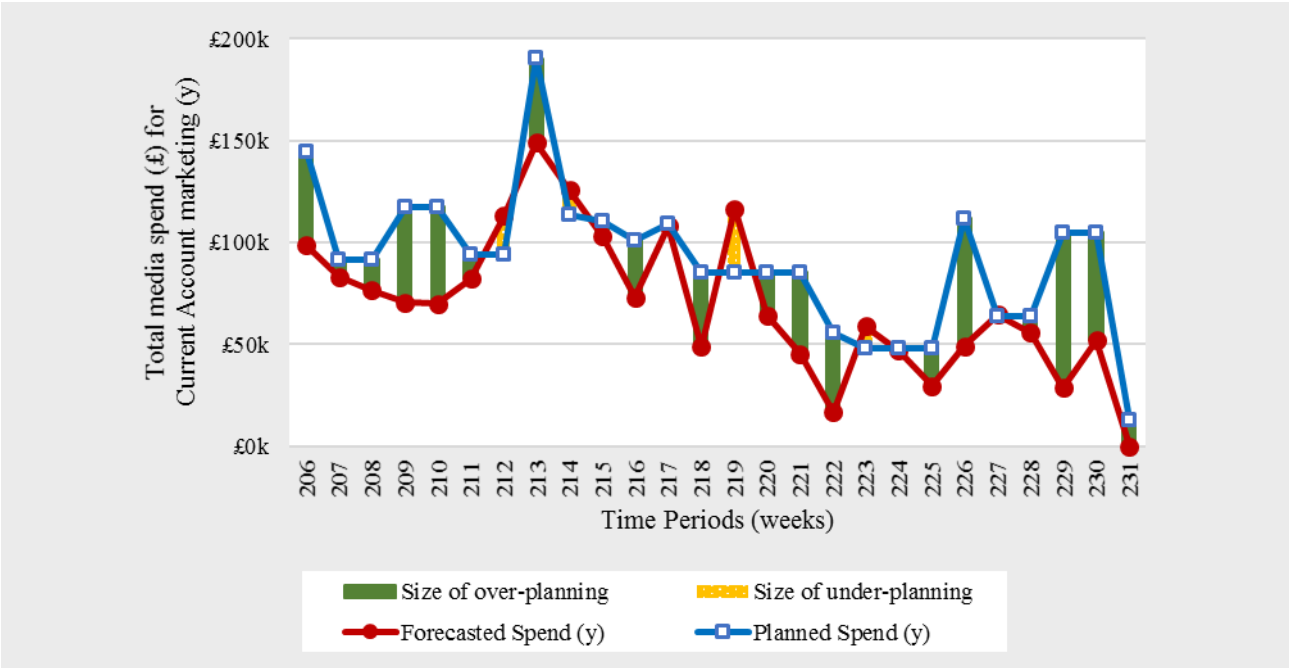
[142] ISA Season, Student Current Account Season and Bond Season binary variables. The models are completed with stepwise linear regression (OLS) where the entry and exit criteria are 0.05 p values.

The combined analysis of the 10 product models based on the Standardised Estimates enables analysts and financial professionals to determine the key factors (and their weights) of the marketing campaign success from a management control perspective. This full analysis fully confirms Hypothesis 4.2.

4.5 EVALUATION OF THE NEW METHODOLOGY

The final step is the evaluation of the new methodology. For this author is still using the predicted cost values from Model 4 (Chapter 4.2.5) and the last dataset collected and desensitised from Bank A. This dataset is based on the cost plan schedule submitted at the end of the year (2015) for the next year (2016). The marketing cost planning schedule is detailing the expected (and provisioned) costs split by product and media type. This plan is based on a combination of new customer volume recruitment target; planned weekly campaign activity to achieve this target; and the cost of the same marketing activity from previous year adjusted by the inflation.

When the observed (or actual) cost is compared against the planned cost (please see graph in Appendix 13) we can see that the forward looking planning with the above detailed planning practice dominantly overestimates the cost (for 18 weeks out of the additional 26) on average by 98% and under estimates in other cases (8 out of 26 periods) on average by 90%, which for the assessed period of time (between weeks 206 and 231) is resulting a total difference of 34% (plus £600k) in the planned cost compared to actual spend. Assuming that the new customer acquisition target is not changing, the main source of difference is the size of the campaign, which is evaluated only a couple of weeks before the actual campaign when more information (e.g. economic environment, market, competitiveness, etc.) is considered by the marketing planning analyst. Therefore, an alternative planning method that predicts cost with a smaller standard deviation to the actuals compared to the plan can improve the accuracy of the planning process. Figure 34 shows the forecasted planned costs with the size of over-planning (green) or the size of under-planning (dotted orange) at each time data points.



Source: Compiled by author

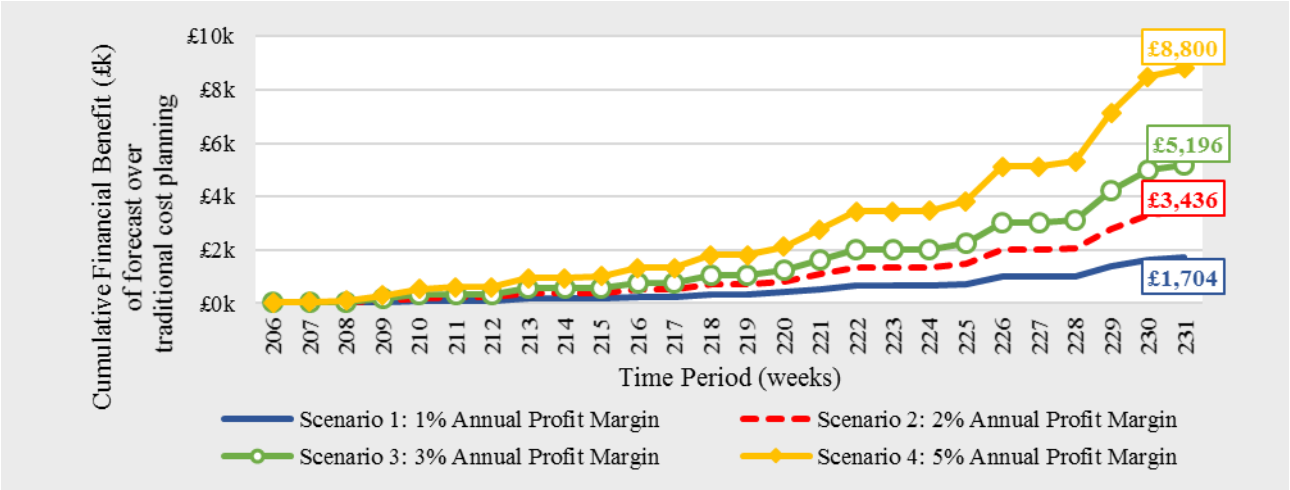
Figure 34: Planned and Forecasted Current Account Media spend

In Chapter 4.2.7[143] author has established that the correlation between the actual spend and the forecasted spend is over 75% and from Appendix 13 we can see that the total difference between the two frequencies is only 2.7%. Using the MAPE ratio[144], the metric author has used for measuring the size of the error in percentage terms to test the customer volume predicting capability of the model, we can statistically prove that the difference between the forecasted and actual values are lower than the difference between the planned and observed spend over the assessed period:

- The MAPE ratio 38.5% for the forecasted values against the actual model;
- The same ratio is 67.23% for the planned spend model.

This proves (that there is) a 28.7%[145] uplift in prediction accuracy per observation and 30.73%[146] in the uplift in the total accuracy, which confirms Hypothesis 5.2 as the campaign cost values predicted by the multivariate linear model are closer to the actual values compared to the campaign cost calculated with existing marketing control planning methods over the actual values. The increased accuracy has multiple impact:

- **Time savings:** marketing planning analysts or financial controllers do not have to constantly re-evaluate the plan but only update the predictors of the model and re-calculate the expected cost.
- **Contractual benefits:** the marketing department can use the calculation for the negotiations to determine the optimal contracting position and negotiation strategy.
- **Financial benefits:** the financial organisation can release the fund difference between the original plan and the forecasted total from the accrued annual spend budget, which means this financial resource can be used for core activity. In order to quantify the financial impact, author has used a simple weekly compound interest calculation with the above described differences. For the annual interest rate the ROI or profit margin figures from the profit and loss accounts can be used (Calculation details and methodology can be found in Appendix 14.). Figure 35 shows the result of the cumulative weekly benefit. The application of the cost forecasting methodology detailed in current research can result 0.5% savings of the annual marketing cost budget.



Source: Compiled by author

Figure 35: Financial Benefit of the new forecast method over traditional planning

[143] Stage VII. – Cost prediction and intervals
[144] As detailed in chapter 4.2.6 Stage VI. – Cross-Validation the model
[145] The difference between the MAPE ratio of the forecasted model (38.5%) and the planned spend model (67.23%)
[146] The total observed cost in the simulation dataset is £1.78m, the total planned is £2.38m which is 33.46% higher than the observed value, the total forecasted value is £1.83m which is only 2.73% higher therefore the total planned cost is 30.73% more than the total forecasted value compared to the observed spend.

4.6 NEW SCIENTIFIC FINDINGS

I. *Author*, based on the reviewed methodologies of existing management control cost planning (especially marketing budgeting) and evaluation procedures, *has found that these models often use multiple market drivers* (base rate, growth of market size, etc.) *in parallel without understanding the interrelations between the different variables* and the underlying factors. *The planning process is often heuristic* based on simplified models of several market and internal factors or using planning methods like: base or zero basis planning, rolling planning, etc.

II. *Current research*, using the desensitised data of an existing large retail bank in the United Kingdom, *has proven that a linear mathematical function or model exists between key bank marketing goals* (such as the number of new customers acquired/recruited or the market share per campaign) *and campaign costs with other factors*.

III. *Author has established the end-to-end procedures, or with other words, the algorithm* (unique sequence of statistical methods and decisions) *to support financial analysts and controllers to conduct marketing campaign cost prediction and evaluation.* Author has processed the literature of other multivariate methods (principal component, factor and regression analysis) and established the main stages of (campaign) cost prediction modelling through the design of simple process flows. *Furthermore, it was also confirmed that multivariate linear regression is an adequate method for marketing campaign cost prediction* with the above detailed marketing goals as the campaign goal predictive equation can be re-used to calculate forecasted cost levels if the goal variable is replaced with targeted or budgeted volumes.

IV. *Author has found evidence for that new cost predictive management control method improves planning accuracy* (on average 28.7% uplift in prediction accuracy per observation and 30.73% in the uplift in the total accuracy when the campaign cost values predicted by the multivariate linear model are compared to the campaign cost calculated with the existing marketing control planning methods and to the actual values). This can be translated to time savings, contractual and financial benefits. Author has estimated that the application of the cost forecasting methodology detailed in current research can result 0.5% savings of the total annual marketing cost budget.

V. Furthermore, using the modelling equation author *has identified the competitiveness of the product proposition variable group as the most influential predictors followed by the new business market share of other products* (competitiveness of the product portfolio) and cross sales activity. Although there is a significant variation on a product group level, *the fourth predictor group is the marketing spend by the financial organisation*. T*he rest of the variance in the assessed models were influenced by seasonality and the economic factors*.

Although the statistical methods in the algorithm are not new and financial organisations are already using them in many ways in their retail customer service, detection of fraud, and improvement of operational efficiencies, the application in the operative planning and evaluation processes of management control alongside with the traditional methods is considered as a new improvement.

4.7 FULFILLMENT OF RESEARCH HYPOTHESES

Based on the comprehensive review of the relevant literature and the completed hypothesis testing, the results were the followings:

Hypothesis 1 (H1)
Existing cost planning, especially marketing campaign cost planning, methods are dominantly non-statistics based approaches. Most (over 75%) of the applied planning (especially cost planning) and evaluation methods described in the selected literature are non-statistical approaches (do not apply multivariate statistical methods for example).

Status: Partially confirmed (Note: Although author has found evidence to prove that literature is dominated by non-statistical approaches, due to the sample size and the selection method (total population unknown, not random sampling, etc.) author acknowledges the result as partial confirmation of the hypothesis.)

Hypothesis 2 (H2)
The field of management control systems, harnessing the big data technology and data science is new and has only been emerging from the literature in the last decade. The literature of management control harnessing big data technology and data science is still developing (it shows an upwards trend with no evidence of a plateau yet).

Status: Confirmed (Note: the combined literature of management control and data science is still considered new and therefore new researches such as current research, would have more contribution and impact when compared against a mature (well-researched and published) field.)

Hypothesis 3 (H3)

Hypothesis 3.1 (H3.1)
In order to provide a statistic based and potentially more accurate and efficient campaign cost planning and evaluation method, a mathematical model between the campaign goals, the marketing costs and other contributing factors must exist. *A mathematical function (model) exists between key bank marketing goals (such as the number of new customers acquired or recruited or market share per campaign) and campaign costs with other factors.*

Status: Confirmed

Hypothesis 3.2 (H3.2)
If such model exists then the start-to-end process to support analysts and controllers to conduct marketing campaign cost prediction and evaluation analysis should be captured. *The statistical process from the conceptualisation of the modelling aim to the statistically validated regression equation, can be captured on simple process diagrams to support financial professionals.*

Status: Confirmed (Note: The start-to-end statistical process was captured on seven process maps to help management control professionals to build time series based multivariate regressive models that can be used for cost prediction or for the analysis of the deviation from the plan. The modelling exercise was completed based on the process diagrams. As author was able to produce a statistically significant regression model it has proven that the process maps contain adequate amount of information to carry out this method successfully)

Hypothesis 4 (H4)

Hypothesis 4.1 (H4.1)
There is a linear relationship between the campaign goals, the campaign spend and most of the other contributing factors (therefore multivariate linear regression is an adequate method).

Status: Confirmed

Hypothesis 4.2 (H4.2)
Using the modelling database and the multivariate linear model, the key factors that determine success of a bank marketing campaign (such as number of new customers and size of market share per campaign) can identified with their weight (or contribution) to the success.

Status: Confirmed

Hypothesis 5 (H5)

Hypothesis 5.1 (H5.1)
The linear equation extracted with the multivariate linear model can be used for cost forecasting and identifying required cost levels for new customer recruitment targets in the system of other contributing factors for future campaigns. *The cost values forecasted by the model for new observations are statistically significant (with other words, the correlation between the forecasted and actual spend values is over 65% and the variances of the forecasted and observed populations are equal at the 95% confidence level).*

Status: Confirmed

Hypothesis 5.2 (H5.2)
The final assessment is the uplift in prediction accuracy. *The campaign cost values predicted by the multivariate linear model are closer to the actual values compared to the campaign cost calculated with existing marketing control planning methods versus actual values.*

Status: Confirmed (Note: Author has found evidence for an average 28.7% uplift in prediction accuracy per observation and 30.73% in the uplift in the total accuracy when the campaign cost values predicted by the multivariate linear model are compared to the campaign cost calculated with the existing marketing control planning methods and to actual values.)

5. CONCLUSION AND DISCUSSION

The results of current dissertation enable the author to draw complex, logical conclusions in the emerging field of management control harnessing data science and big data capabilities.

Current Planning Practice: The extensive review of the Hungarian, German and British-American literature highlighted the importance of the planning and evaluation (as part of supervision) functions within management control. Despite of the insufficient number of publications for complete representability, it can be established that *the planning and evaluation methods* described in practical and scientific works *are dominantly heuristic methods*. These are primarily based on the planning practice and experience of previous planning cycles (e.g. base budgeting, zero base budgeting, etc.) as well as based on estimated activities (e.g. rolling forecast). These approaches are simple however are limited as they focus on one main driver in general.

Less than 40% of the reviewed literature described at least one analytical or statistical approach, however other methods described in these publications are still dominantly non-statistical based methods. The described analytical campaign budgeting processes mainly focus on the probability of events and in some cases on the correlation between the communication budget and a target value applying basic statistical procedures. With the fast-paced development of big data technologies author identified an emerging new field; the data science and big data technology applying management control, which offers a range of statistical solutions for traditionally accountancy and heuristic questions, such as planning and evaluation, with speed and efficiency.

Application of data science and big data technologies: There are several ways for financial organisations to gain business advantages through data mining and analysing. Banks already use advanced statistical models for example, as part of their Customer Relationship Management (CRM) activity banks build propensity models[147] for direct mailing/phone campaigns based on their structured and extensive databases, banks also use predictive techniques (decision trees, neural networks, etc.) to detect fraud or to size the risk of default for loan products. But big data can also be used to identify exposure in real time across a range of sophisticated financial instruments such as derivatives or foreign currency exchange. Predictive analysis of large internal and external data can result better risk management, improved customer loyalty and ultimately uplift in profitability.

Author identifies the *improvement of operational efficiencies with big data and data science in the field of management control as another, new way of application*. Both topics of management control and data science are extensively researched and wide ranges of publications are available[148], which provide firm bases for these fields. At the same time, the number of combined researches of the management control and data science is relatively low[149], which suggest that there is an opportunity to improve existing operations, processes or methods. The assessment of the other, potential applications in the field of management control is a potential research topic not covered by current dissertation. The application of machine learning can potentially automate several planning and evaluation activities, which will eventually have an impact on the performance and capabilities of existing management systems. Furthermore, such technology will also require its own specialised

[147] A model predicting a customer's potential to perform an action, e.g. apply for a current account, etc.
[148] Management control: over 36 thousand publications since 1960, Data Science: over 18 thousand publications since 1985 in the ProQuest publication database.
[149] Less than 260 publications since 2000 in the ProQuest publication database.

management control area either as part of the information technology management control or as a big data technology management control area. This will possibly require the further specialisation of the management control profession, which is another potential research topic for the future. Author at this stage expects management control professionals with primarily financial, accountancy and secondarily statistical and computer science background to utilise the new proposed model therefore author has chosen a simpler statistical algorithm (a process to be followed in the problem-solving, planning operation) from the toolset of data science, i.e. regression modelling combined with dimension reduction procedures[150].

Regression modelling (opportunities and pitfalls): Regression modelling can be complex to apply and interpret depending on the order of the model therefore author's aim was to prove that even a first order (linear) regression model can be effectively and accurately applied to predict new customer volumes and ultimately to forecast cost volumes. Furthermore, to confirm that the new data science (statistics) based marketing campaign cost planning method can be captured on simple process diagrams. It can be argued if the linear relationship is always appropriate to describe the interrelations of variables and whether forcing the model to explain the variance of the target variable only by variables in linear relationship is not causing the loss of meaning or explanatory power. Therefor a further research on simple, automated, non-linear methods is required in the future.

Author wanted to test the assumption whether a linear relationship between the new customer requirement volumes and the marketing campaign cost in the web of other contributing factors is adequate and applicable. The research, using the desensitised data of an existing large retail bank in the United Kingdom, has proven that *a linear mathematical function or model exists between key bank* marketing goals (such as the number of new customers acquired or recruited or market share per campaign) *and campaign costs with other factors*. Author has confirmed that *multivariate linear regression is an adequate method for marketing campaign cost prediction* with the above detailed marketing goals as the campaign goal predictive equation can be re-used to calculate forecasted cost levels if the goal is replaced with targeted or budgeted volumes. Furthermore, *the statistical process, from the conceptualisation of the modelling aim to the statistically validated regression equation, can be captured on simple process diagrams to support financial professionals* as the valid regression model was built using the Process Diagram of Stage I.-VII.

Although linear regression modelling offers a simple and efficient way to predict marketing campaign cost, it is important to highlight that time-series based linear regression requires increased attention from the analyst as such models have a potential to violate the basic assumption of linear regression modelling due to the serial correlation between the consecutive records (or observations). Once the autoregressive errors are treated appropriately (as described in the research) the method offers a range of opportunities and improved efficiency.

The impact of the new method: *The new algorithm can be applied for planning to predict campaign goals or costs* directly as well as to forecast cost indirectly based on campaign goal modelling equation. These equations *can also be used for cost evaluation to determine whether a given marketing campaign spend is optimal based on the historic data*. With the use of DFBETAS function *the analyst can understand the separate impacts of the contributing variables* (e.g. marketing spend)

[150] Principal Component Analysis, Factor Analysis

period by period (e.g. weeks, months), which provides a deeper insight into the comparison analysis of budgeted figures and actuals.

Author has *found evidence for an average of 28.7% uplift in prediction accuracy per observation and 30.73% in the uplift in the total accuracy* when the campaign cost values predicted by the multivariate linear model are compared to the campaign cost calculated with the existing marketing control planning methods and to actual values. The increased accuracy has multiple impacts: time savings (the marketing planning analyst or the financial controller do not have to constantly re-evaluate the plan but only update the predictors of the model and re-calculate the expected cost); contractual benefits (the marketing department can use the calculation for the negotiations to determine the optimal contracting position and negotiation strategy); financial benefits (the financial organisation can release the fund difference between the original plan and the forecasted total from the accrued annual spend budget which means this financial resource can be used for core activity. Author has estimated that the application of the cost forecasting methodology detailed in current research can result 0.5% savings of the total annual marketing cost budget).

The application of data science in management control can be further improved with the automation of the above elaborated approach. Further development option would be the linkage of real time internal and external data through APIs. This would enable constant computer based cost evaluation and forecasting as well as the instant determination of customer profitability or the changes of the market, which eventually improve the options of the management. The model equation can also be used to determine key strategic focus and ultimately to promote a better (campaign goal oriented) distribution of resources of across contributing factors.

The key success factors of new banking customer base growth: Based on the initial variables identified by the focus group of industry professionals and the subsequent modelling exercise, the *competitiveness of the product proposition has been identified as the highest influencer variable group*. The *second group is the new business market share of other products,* which highlights the importance of a strong product portfolio of the financial organisations. Strong campaign performance (or high new business market share) of other products such as personal loans and mortgages are highly advertised and in case of competitive rates are also promoted by intermediaries, which ultimately increases brand and product awareness and can be translated to actual product openings. This variable group should also be analysed with the *cross sales activity* as a significant part (15% for current accounts) of the new current account customers are recruited in junction with another product. *The fourth predictor group is marketing spend*. There is a significant variation on a product group level as the marketing spend has the highest impact on the Savings campaigns (17%) and the lowest on Credit Cards (5%). The rest of the variance in the assessed models were influenced by seasonality and the economic factors.

Further developments: Author suggests: *1.)* the assessment of the application of machine learning (to automate several planning and evaluation activities; *2.)* the establishment of a new specialised management control area (big data and information technology management control); *3.)* the review of the controller role (tasks responsibilities, required new skills and education); *4.)* expansion of the research horizontally (repeat the current research with other initial variables and also with modelling data sourced from other or multiple financial organisations) and vertically (apply he algorithm for other cost groups; *5.)* develop the methodology further (application of advanced methods like neural networks and decision trees etc.).

6. SUMMARY

The bank sector has a vital role in the modern economy. It is a sensitive monitor, engine and control factor of the economic system. Following the recent financial crisis, the focus is on re-regulation, the measurements of risk, efficiency and profitability as well as on value adding processes for the investors and the customers. In order to achieve this, banks – among other structural and cultural changes - have to develop effective planning and controlling systems adapted to the new challenges and focusing on financials, customers and market at the same time. Based on empirical experience author has identified a potential improvement area with the combination of statistics, data science and management control enabled by the fast and efficient big data technology.

Author in current research aimed to establish an alternative procedure that supports marketing cost planning and the post campaign evaluation (actual versus budget analysis) for primarily current account banking products (with also credit card and savings products taken into account). There were two campaign aims considered for the model building and post campaign analysis: increase the number of customers by the targeted volume; and reach targeted market campaign share. It is important to highlight that the proposed procedure is generalised so it can be used for other campaign aims as well however other goals did not form part of the current research scope. The simplicity and efficiency of the new procedure were two vital aspects as the improvement was intended to be used primarily by financial controllers and only secondarily by statisticians. Therefore, author was aiming to apply the simple method of multivariate linear regression with simple process maps. The application of non-linear regression modelling methods was out of scope of current dissertation.

In order to lay down firm foundations for the research, author has identified the key British-American, German and Hungarian authors to review existing literature of controlling and management control: starting from the definition of the controlling concept; through the exploration of the evolution of management control; to the review of the role of controlling in the modern economy, specifically the interactions with the regulatory environment. Furthermore, author has outlined the system, aims, functions, tools and different levels of management control at different organisations (detailing the specifications of controlling at financial institutions). Similar review has been carried out for the marketing specific management control area where the relevant marketing management theories, concept and functions; the interrelation between marketing management and the controlling concept; and the marketing management tools with the strategic and operative tasks have been systematically assessed. Author has also focused on the literature of management control planning and supervision procedures and methods, especially on cost planning, marketing budgeting and evaluation. Finally, author has reviewed the enablers of the current methodology improvements; the data science and big data technology.

As part of the methodology review, author has established the end-to-end procedures or with other word, the algorithm (unique sequence of statistical methods and decisions) to support financial analysts and controllers to conduct marketing campaign cost prediction and evaluation. Author has processed the literature of other multivariate methods (principal component, factor and regression analysis) and established the main stages of (campaign) cost prediction modelling through the design of simple process flows.

In the results section, based on the reviewed methodologies of existing management control cost planning (especially marketing budgeting) and evaluation procedures, author has categorised the planning methods by the level of mathematics and statistics involved. Author has found that these models often use multiple market drivers (base rate, growth of market size, etc.) without understanding the interrelations between the different variables and the underlying factors. The planning process is often heuristic, based on simplified models of several market and internal factors or using planning methods like: base or zero basis planning, rolling planning, etc.

Based on the above detailed findings and the review of the latest developments of data science in the financial industry, author identified a field of management control to improve operational efficiencies with big data and data science through a new cost planning method. At the end of the process of defining the concept of big data and data science; and reviewing the existing and potential applications of data science in the financial industry (e.g. campaign cost prediction, etc.), author established that both topics of management control and data science are extensively researched and a wide range of publications is available, which provides firm basis for these fields. At the same time, the number of combined researches of the management control and data science is relatively low, which suggests an opportunity to improve existing operations, processes or methods.

The research, using the desensitised data of an existing large retail bank in the United Kingdom, has proven that a linear mathematical function or model exists between key bank marketing goals (such as the number of new customers acquired/recruited or the market share per campaign) and campaign costs with other factors. Author has confirmed that multivariate linear regression is an adequate method for marketing campaign cost prediction with the above detailed marketing goals as the campaign goal predictive equation can be re-used to calculate forecasted cost levels if the goal variable is replaced with targeted or budgeted volumes.

Author has also proven that new cost predictive management control method improves planning accuracy (on average 28.7% uplift in prediction accuracy per observation and 30.73% in the uplift in the total accuracy when the campaign cost values predicted by the multivariate linear model are compared to the campaign cost calculated with the existing marketing control planning methods and to actual values). This can be translated to time savings, contractual and financial benefits (author has estimated that the application of the cost forecasting methodology detailed in the current research can result 0.5% savings of the total annual marketing cost budget).

Furthermore, using the modelling equation, analysts can determine the weight and importance of factors contributing to the success of marketing campaigns, which ultimately can help with the optimal allocation of resources. Author has identified the competitiveness of the product proposition variable group as the most influential predictors followed by the new business market share of other products (competitiveness of the product portfolio) and the cross sales activity. Although there is a significant variation on a product group level, the fourth predictor group is the marketing spend by the financial organisation. The rest of the variance in the assessed models were influenced by seasonality and the economic factors.

Although the statistical methods in the algorithm are not new and financial organisations are already using them in many ways in their retail customer service, detection of fraud, and improvement of operational efficiencies, the application in the operative planning and evaluation processes of management control alongside with the traditional methods is considered as a new improvement.

7. ÖSSZEFOGLALÁS

A bankszektor jelentős szerepet tölt be a modern gazdaságokban. Egyszerre működik a gazdaság érzékeny jelzőrendszere, motorja, illetve szabályozójaként. A pénzügyi válságot követően a figyelem középpontjában az újra-szabályozás, a kockázatok és a hatékonyság mérése, a jövedelmezőség javítása került a tulajdonosok és az ügyfelek számára történő értékteremtés segítségével. Annak érdekében tehát, hogy a bankok a megváltozott környezetben is hatékonyan tudjanak működni, egy stabil kontrolling rendszerre van szükségük, amely képes az új kihívások és technológiai fejlődés eredményeit hatékonyan alkalmazni. Empirikus tapasztalataim segítségével a kontrolling tudományág egy új feltörekvő területét, a „Big Data" technológia és az adattudomány eszköztárát alkalmazó menedzsment kontrollt azonosítottam, amely a tradicionálisan heurisztikus módszerek helyett gyors, pontos és hatékony statisztikai alapú tervezési és értékelési megoldásokat kínál.

A kutatás keretében célul tűztem ki a menedzsment kontrol által alkalmazott marketing költségtervezési és értékelési módszereket vizsgálatát, valamint az alkalmazott módszertan fejlesztési lehetőségeit statisztikai módszerekkel és az adattudomány eszköztárával. Az elsődleges célom egy új, az eddigi módszertant kiegészítő tervezési és értékelési eljárás statisztikai folyamatának, algoritmusának (folyamatok, eljárások és döntések egyedi kombinációjának) a kidolgozása volt. Az új folyamattal egyrészről a banki kontrollerek és pénzügyi szakemberek folyószámla, hitelkártya és megtakarítási számla termékkel kapcsolatos bank-marketing kampányköltség tervezési és ellenőrzési feladatait kívántam támogatni. Más részről a kialakított modellek alapján meg kívántam határozni a fent említett banki termék-kampányok sikertényezőt és ezen tényezők dinamikusan változó összetételét, ezzel is segítve a bank kontrolling tervezési és elemzési funkcióját.

A kutatás szakmai megalapozásának érdekében a magyar, német, angol és amerikai szakirodalom széleskörű áttekintését végeztem el. A szakirodalmi feldolgozás részeként megvizsgáltam a menedzsment kontrol és a kontrolling szakirodalmát, meghatároztam a kontrolling koncepcióját, valamint annak céljait, funkcióit és rendszerét kitérve a hitelintézeti kontrolling sajátosságainak feltárására. Részletesen vizsgáltam az általános kontrollingon belül található marketing kontrolling funkcionális területét, továbbá a kontrolling stratégiai és operatív tervezési, valamint ellenőrzési és irányítási funkcióit. A szakirodalmi áttekintést a fejlesztést elősegítő adattudomány és a „Big Data" technológia vizsgálatával zártam, kitérve a téma relevanciájának vizsgálatára, a létező hitelintézeti alkalmazásokra, valamint a marketing kontrollinggal kapcsolatos alkalmazási lehetőségek (például kampányköltség előrejelzés) vizsgálatára.

Az anyag és módszertan részben bemutatásra került a kutatási stratégia, különösképpen a kutatási célok, hipotézisek, az alkalmazott primer (kvalitatív kutatás) és szekunder adatforrások (külső és belső banki adatbázisok), valamint a hipotézisvizsgálat módszertanának rendszerszerű vizsgálata. Továbbá sor került az új preditkív költségtervezési modell részletes statisztikai, valamint grafikus (folyamatábrákkal történő) ismertetésére a főbb matematikai ismeretek (például a faktor, főkomponens és korreláció elemzés, többváltozós lineáris regresszió elemzés autoregresszív hibával, stb.) és döntési javaslatok (például a minimális input adatok mennyisége és részletezettsége) aprólékos bemutatásával együtt, ezzel segítve a kontrollereket és pénzügyi szakembereket a marketing kampányköltség tervezési folyamatokban (mint egy használati útmutató).

A hipotézisvizsgálatok és elemzések az Eredmények című fejezetben kerültek bemutatásra. Elsőként a jelenleg alkalmazott költség tervezési (különös tekintettel a marketing kampányköltség) és értékelési kontrolling eljárások szakirodalmának rendszerszemléletű áttekintése alapján megállapítottam, hogy a jelenlegi módszerek jellemzően heurisztikus eljárások, azaz az előző év vagy tervezési időszak tény értékéből indulnak ki és egyszerű extrapolációs módszer segítségével határozzák meg a következő időszak költségtervét (inflációs korrekció, bázis tervezés, 0 bázisú vagy gördülő tervezés). Számos modell használ például olyan ún. drivereket, amelyek egymással párhuzamosan, a kölcsönöshatásokat figyelmen kívül hagyva működnek a látens tényezők feltárása nélkül. Ezt követően a fejlesztés alapfeltételeit vizsgáltam. Megállapítottam, hogy a menedzsment kontrol és a módszertani fejlesztéshez szükséges adattudomány kiterjedten kutatott és publikált tudományágak, amelyek megfelelő alapot szolgáltatnak a fejlesztésre. Ugyanakkor, az adattudomány és a „Big Data" technológiát alkalmazó menedzsment kontrol kombinált tudománya újnak tekinthető (az utóbbi évtizedben kezdett kialakulni), így ezen a területen végzett kutatások hatása nagyobb, mint más, kiterjedten kutatott tudományágak esetében.

Az angliai kereskedelmi banktól származó, szenzitív vállalati információtól tisztított adatokat, terveket és célokat tömörítő adatbázis segítségével bizonyítottam, hogy egy lineáris függvényszerű összefüggés írható fel a főbb bank marketing célok (új ügyfélszám növelése vagy piaci részesedés adott kampányban) és a kampányráfordítások, valamint egyéb befolyásoló tényezők között. A bizonyításhoz szükséges matematikai modellezés során sikeresen kialakítottam az új kampányköltség tervezési eljárás statisztikai folyamatát, algoritmusát a kezdeti tervezési célok meghatározásától a hitelesített regressziós egyenlet kidolgozásáig és alkalmazásáig, egyszerű folyamatábrák segítségével.

Ezt követően bizonyítottam, hogy a statisztikai alapon készített, új prediktív költségtervezés javítja a tervezés pontosságát és hatékonyságát. (A statisztikai alapon készített költségterv esetében megfigyelésenként átlagosan 28.7%-os javulás figyelhető meg a pontosság tekintetében, míg a teljes időszaki előrejelzett költség 30.73%–al pontosabb a hagyományos (heurisztikus) módszerrel készített költségtervhez képest a megfigyelt vagy tényleges költségekhez viszonyítva.) Továbbá a pontosabb költségtervnek számos járulékos előnyét azonosítottam: időmegtakarítás a gyorsabb és ritkább újra tervezés révén, szerződéses előnyök az információval támogatott tárgyalási pozíció alapján, valamint pénzügyi előnyök a le nem kötött, valamint el nem határolt pénzügyi források miatt.

A lineáris egyenlet elemzése során arra a megállapításra jutottam, hogy a termék versenyképességét mérő változók csoportja van a legjelentősebb hatással az új banki ügyfelek számának növekedésére, ezt követi a többi termék kampány időszaki piaci részesedése (azaz a portfólió versenyképessége), illetve keresztértékesítés mértéke. Habár jelentős különbségeket tártam fel a különböző termékcsoportoknál, a negyedik tényező csoport a marketing ráfordítás és az aktivitás. A differenciált sztenderd becslések (Béták) segítségével matematikailag (és grafikusan) kimutattam a fent említett tényezők dinamikusan, idő-periódusonként (hetenként) változó hatását és összetételét.

A kutatás során alkalmazott statisztikai és matematikai módszertani elemek széleskörben elterjedt eljárások, ugyanakkor a banki gyakorlati felhasználás az ügyfél- és kockázatelemzés, valamint csaláselhárítás területeire összpontosít. A kutatás újdonság tartalma, az adattudomány eszközeinek egy tradicionálisan számviteli eljárásokkal dominált funkcionális kontrolling területén, a marketing menedzsment kontrol területén történő alkalmazásában nyilvánul meg, mintegy kiegészítve a meglévő módszertani elemeket.

APPENDIX

A1. References

1. ALGINA, J.; OLEJNIK, S. (2000): Determining sample size for accurate estimation of the squared multiple correlation coefficient. *Multivariate Behavioral Research*, Vol (35), No (1), pp. 119-136. ISSN: 0027-3171
2. ALLAIRE, Y.; FIRSIROTU, M.E. (1984): Theories of organizational culture. *Organizational Studies*, Vol (5), No (3), pp. 193-226 ISSN: 0170-8406
3. AMAT, J.M. (1992): Management Accounting Systems in Spanish Firms. *The European Accounting Review*. Vol (1), No (1), pp. 1-26. ISSN: 0963-8180
4. ANTHONY, R.N. (1965): Planning and Control Systems: a Framework for Analysis. *Studies in Management Control*. Division of Research, Harvard University Graduate Business School of Business Administration, ISBN: 0-87584-0477
5. ANTHONY, R.N. (1988): The Management Control Function – 5th Edition. Harvard Business School Press, USA. pp. 25-60. ISBN: 978-0-8758-4184-7
6. ANTHONY, R.N.; DEARDEN, J. (1980): Management Control Systems (IV. Edition), The Robert N. Anthony/Willard J. Graham series in accounting R.D. Irwin, Homewood, Illinois. ISBN: 0256-02325-5
7. ARGYRIS, C. (1990): The dilemma of implementing controls: the case of managerial accounting. *Accounting, Organization and Society*, Vol (15), No (6), pp. 503-511 ISSN: 0361-3682
8. ARMITAGE, P.; BERRY, G.; MATTHEWS, J.N.S. (2001): Statistical Methods in Medical Research. Fourth Edition. Blackwell Publishing Company, Oxford, pp. 312-377. ISBN: 0-632-05257-0
9. ARLOT, S.; CELISSE, A. (2010): A survey of cross-validation procedures for model selection. Statistical Surveys. Vol (4) pp. 40-79. ISSN: 1935-7516 Accessed on 23/10/2016 21:00 Downloaded from: http://projecteuclid.org/download/pdfview_1/euclid.ssu/1268143839
10. BAGINÉ, D.E.; KOLONICS, R.; SZLOBODA, I. (1999): Gyakorlati controlling 1-3. Magyarországi vállalkozások és intézmények controlling kézikönyve, Weka Szakkiadó Kft., Budapest, pp.: 1-8.; 19-25. ISBN: 963-859-205-2
11. BARNARD, C.I. (1938): The Functions of the Executive, Cambridge, USA MA Harvard University press, pp. 8 ISBN: 9-780-6743280-06
12. BARRETT, R.; HAWTHORN, L. (2005): The 123 of ABC Methodologies – time-splits, time-capture & time-driven, Armstrong Laing Group, Chicago, Last accessed on 03/03/2017 22:10 from costkiller.net/tribune/Tribu-PDF/The-123-of-ABC-Methodologies.pdf
13. BARTLETT M.S. (1937): Properties of Sufficiency and Statistical Test. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences Vol (160), No (901), pp. 268-282. Royal Society, London. Accessed on 26/09/2016 from http://www.jstor.org/stable/96803
14. BEEL, J.; LANGER, S. (2011): An Exploratory Analysis of Mind Maps. In Proceedings of the 11th ACM Symposium on Document Engineering, Mountain View, California, USA, pp.: 81-84 2011. ACM. Downloaded from www.docear.org
15. BENGIO, Y.; GRANDVALET, Y. (2003): No Unbiased Estimator of the Variance of K-fold Cross-Validation. *Journal of Machine Learning Research*. Vol (5) pp. 1089-1105 ISSN: 1533-7928 Downloaded from: http://www.jmlr.org/papers/volume5/grandvalet04a/grandvalet04a.pdf Accessed on 23/10/2016
16. BERRY, A.J.; BROADBENT, J.; OTLEY, D. (1995): Management control: theories, issues and practices, London, UK. MacMillan Limited Press. pp. 3-26; 283-298 ISBN: 978-0-3335-7243-6

17. BIRNBERG, J.G. (1999): Management Accounting Practice and research as we end the twentieth century. IN: EPSTEIN, M.J.; LEE, J.Y. ed. Advances in Management Accounting. Vol (8), Bingley: Emerald Group Publishing Limited. pp.: 1-27. ISBN: 978-0-7623-0620-6

18. BIRNBERG, J.G. (2000): The role of behavioural research in management accounting in the 21st century. *Issues in Accounting Education.* Vol (15), No (4), pp. 713-728. ISSN: 0739-3172

19. BODA, GY.; CSANÁDI, P.; PÁL, G. (2001): A kontrolling, mint a menedzsment technikák párbeszéde. *Kontrolling*, 2001, Vol (1), pp. 2-5.

20. BODA, GY.; SZLÁVIK, P. (2001): Vezetői controlling. Közgazdasági és Jogi Kiadó, Budapest.BOX, E.P.G.; JENKINS, G.M.; REINSEL C.G. (2008): Time series analysis: Forcasting and Control. Fourth edition. New Jersey, USA. John Wiley & Sons, Inc. pp. 93-136., ISBN: 978-0-470-27284-8

21. BURTON, R.M.; OBEL B. (2013): Design rules for dynamic organization design: the contribution of computational modelling. IN: GRANDORI, A. ed. Handbook of Economic Organization: Integrating Economic and Organization Theory, Edward Elgar Publishing Limited, Massachusetts, USA, pp.: 223-244. ISBN: 978-1-84980-398-4

22. BÖCSKEI, E. (2006): A marketingcontrolling felértékelődése az információs iparágak piacán. *A Controller*, Vol (2), No (11), pp. 13-15. ISSN 2063-9309

23. BRÁVÁCZ, I. (2008): A marketingkutatás szerepe a marketingcontrolling tervezési, eltéréselemzési és információkezelési tevékenységében. *A Controller*, Vol (4), No (4), pp. 22-24. ISSN 2063-9309

24. CAPLAN, E.H. (1971): Management accounting and behavioural science, Massachusetts, USA. Addison Wesley Publishing Company Inc. pp. 5-28. ISBN: 978-0-2010-0891-3

25. CARENYS, J. (2012): Management Control Systems: A Historical Perspective. *International Journal of Economy, Management and Social Sciences.* Vol (1), No (1), December 2012, pp. 1-18.

26. CAVANILLAS, J.M.; CURRY, E.; WAHLSTER, W. (2016): New Horizons for a Data-Driven Economy. A roadmap for usage and exploitation of big data in Europe. Springer Open, USA, pp. 3-11., ISBN 978-3-319-21568-6

27. CHANDLER, A.D. (1977): The Visible Hand: The Managerial Revolution in American Business. Harvard: University press, Sixteenth printing, 2002. pp. 15-185., 209-239., 381-454. ISBN: 0674-940520

28. CHENHALL, R. (2003): Management control system design within its organizational context: Findings from contingency-based research and directions for the future, *Accounting, Organizations and Society*, Vol (28) No (2-3), pp. 127-168. ISSN: 0361-3682

29. CHIKÁN, A. (1997): Vállalatgazdaságtan. AULA Kiadó, Budapest, pp. 9-32., ISBN: 978-9-639698604

30. CHUA, F. (2013): Big Data: its power and perils – Accountancy Futures Academy, The Association of Chartered Certified Accountants (ACCA), London, Last accessed on 28/02/2017 16:52 from: www.accaglobal.com/futures

31. CLARKE, P.D.; GARDENER, E.P.M.; FEENEY, P.; MOLYNEUX P. (1988): The Genesis of Strategic Marketing Control in British Retail Banking. *International Journal of Bank Marketing*, Vol (6), No (2), pp. 5-19 ISSN: 0265-2323

32. CLARK, B.H. (1999): Marketing Performance Measures: History and Interrelationships. Journal of Marketing Management, Vol (15) No (8), pp. 711-732. ISSN:2333-6080

33. COCHRANE, D.; ORCUTT, G.H. (1949): Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*. Vol (44) No (245) pp. 32-61. ISSN:1537-274X Downloaded from: http://dx.doi.org/10.1080/01621459.1949.10483290 Accessed on 06/10/2016

34. COLLINS, F. (1982): Managerial accounting systems and organizational control: a role perspective. *Accounting, Organizations and Society*. Vol (7), No (3) pp. 107-121., ISSN: 0361-3682

35. CONOVER, W.J. (1999): Practical Nonparametric Statistics. Third Edition, John Wiley & Sons Inc., New York, pp. 428-433. ISBN: 978-047-116-06-87

36. CONOVER, W.J.; JOHNSON, M.E.; JOHNSON, M.M. (1981): A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*. Vol (23), No (4), *American Statistical Association and American Society for Quality*, pp. 351-361. ISSN: 0040-1706 Accessed on 30/09/2013 22:38 from: http://www.jstor.org/stable/1268225

37. COOPER, D. (1983): Tidiness, muddle and things: commonalties and divergences in two approaches to management accounting research. *Accounting, Organization and Society*, Vol (8), No (2/3), pp. 269-286. ISSN: 0361-3682

38. CURRY, E. (2016): The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. IN: CAVANILLAS, J.M.; WASHLSTER, W.; CURRY, E. ed. New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe. Springer Open, Switzerland, pp.: 29-37. ISBN: 978-3-319-21569-3 Accessed from: http://link.springer.com/content/pdf/10.1007%2F978-3-319-21569-3.pdf on 30/10/2016 15:53

39. CSATH, M. (2004): Stratégiai tervezés és vezetés a 21. században, Nemzeti Tankönyvkiadó., Budapest, pp. 51-56, 139-151. ISBN: 963-1952-517

40. DERMER, J.D.; LUCAS R.G. (1986): The illusion of managerial control. *Accounting, Organization and Society*, Vol (11), No (6), pp. 471-482. ISSN: 0361-3682

41. DHAR, V. (2013): Data science and prediction. *Communications of the ACM*, Vol (56), No (12), pp. 64-73. ISSN: 0001-0782

42. DILLER, H. (1998): Marketingplanung (2nd edition), Verlag Vahlen, München, pp. 71-74., ISBN:978-3-8006-2271-9

43. DOMÁN, CS. (2005): Többváltozós korreláció- és regressziószámítás. Oktatási Segédlet. Miskolci Egyetem, Miskolc. pp. 3-25.

44. DRAPER, N.R.; SMITH, H. (1998): Applied Regression Analysis – Third Edition. John Wiley & Sons. New York, pp. 47-134. ISBN: 0-471-17082-8

45. DRUCKER, P. (1993): Management: tasks, responsibilities, practices. Harper & Row Publisher Inc., USA, pp. 517-541. ISBN:0-88730-615-2

46. DSSC – DIVISION OF STATISTICS AND SCIENTIFIC COMPUTING (1995): Factor Analysis Using SAS PROC FACTOR. ACITS, The University of Texas at Austing, Statistical Services, 475-9372 Accessed on 09/08/2013 from: http://ssc.utexas.edu/docs/stat53.html

47. DU, Q.L.; CAO, S.M.; BA, L.L.; CHENG, J.M. (2008): Application of PDCA Cycle in the Performance Management System, Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, China, 12-17 October 2008. Accessed on 18/01/2017 from: http://ieeexplore.ieee.org/abstract/document/4679871/

48. FARRIS, P.W.; BENDLE, N.T.; PFEIFER, P.E.; REIBSTEIN, D.J. (2008): Marketingmérések: 50+ mutató, amelyet valamennyi vezetőnek ismernie kell, Scolar Kiadó, Budapest ISBN:978-9-6324-4063-7

49. FAYOL, H. (1949): General and Industrial Management – Reprint (2013), Martino Fine Books, Eastford, USA. pp. 1-52. ISBN: 978-1614274599

50. FARRIS, P.W.; BENDLE, N.T.; PFEIFER, P.E.; REIBSTEIN, D.J. (2008): Marketingmérések (50+ mutató, amelyet valamennyi vezetőnek ismernie kell), Scolar Kiadó, Budapest, pp. 8-22. ISBN: 978-963-244-063-7
51. FIELD, A. (2009): Discovering Statistics Using SPSS: Introducing Statistical Methor (3rd ed.) USA. Sage Publications Ltd. pp. 627-685. ISBN:978-1-84787-907-3
52. FISHER, J. (1995): Contingency-based research on management control systems: categorization by level of complexity. *Journal of Accounting Literature.* Vol (14), pp. 24-53., ISSN: 0737-4607
53. UNITED STATES. (2010): *The financial crisis inquiry report: final report of the National Commission on the Causes of the Financial and Economic Crisis in the United States.* Washington, DC, Financial Crisis Inquiry Commission. http://purl.fdlp.gov/GPO/gpo50165.
54. FLAMHOLTZ, E. (1983): Accounting, budgeting and control systems in their organisational context: theoretical and empirical evidences. *Accounting, Organizations and Society*. Vol(8), No(2) pp. 153-169., ISSN: 0361-3682
55. FRANCSOVICS, A. (2005): A Controlling Fejlődésének Sajátosságai. Doktori (PhD) értekezés. Budapesti Corvinus Egyetem. Budapest. pp. 12-68.
56. FOSTER, G.; GUPTA, M., (1994): Marketing, cost management and management accounting. *Journal of Management Accounting Research,* 6, pp. 43. ISSN: 1049-2127
57. GAÁL, N. (2007): Theoretical Background of Marketing-controlling, Óbudai Egyetem. Budapest, Online publication: http://kgk.uni-obuda.hu/sites/default/files/Gaal_Noemi.pdf (Last downloaded: 12/02/2017)
58. GÁCSI, R.; ZÉMAN, Z. (2013): Marketing Performance Management. *International Journal of Business and Management Studies*. No (6) pp.403-413. ISSN: 1309-8047
59. GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S.T. (2003): The Google File System, 19th ACM Symposium on Operating Systems Principles, Lake George, NY, October, 2003. Last accessed on 06/03/2016 from: https://research.google.com/archive/gfs-sosp2003.pdf
60. GRANGER, C.W.J. (1986): Forecasting in Business and Economics – 2nd edition. Emerald Group Publishing Limited, Bingley, UK, pp. 47-92 .ISBN: 978-0-12-295184-8
61. GROSS, A. L. (1973). How large should a sample size be in a regression analysis? In proceedings of the Annual Convention of the American Psychological Association, Montreal, Quebec, Canada, pp.: 17-18. Downloaded: http://www.apa.org/about/apa/archives/meeting-directory.aspx#1970 on 04/04/2015
62. GYULAINÉ ZS.ZS.; JÁGERNÉ SZ. M. (2006): Banki controlling és belső ellenőrzés a gyakorlatban. Perfekt Gazdasági Tanácsadó, Budapest, pp. 11-72., ISBN: 963-394-640-9
63. HÁGEN, I. ZS. (2008): A kis- és középvállalkozások versenyképességének növelése kontrollinggal. Doktori (PhD) értekezés. Szent István Egyetem. Gödöllő. p.5.
64. HAJDU, O.; PINTÉR, J.; RAPPAI, G.; RÉDEY, K. (1998): Statisztika I. Janus Pannonius Tudományegyetem Közgazdaságtudományi Kar, Pécs, pp. 280-314., ISBN: 963-641-356-8
65. HAJDÚ, N. (2013): Marketingértékelési szemlélet és hazai tapasztalata. Doktor (PhD) értekezés, Miskolci Egyetem, Gazdaságtudományi Kar, Miskolc pp. 56-63. Last accessed on 05/03/2016 08:08 from: 193.6.1.94:9080/JaDoX_Portlets/documents/document_15477_section_11861.pdf
66. HAJDÚ, N. (2016): A marketingcontrolling feladata és eszköztára. *Controller Info*, Vol (4), No (2), pp. 22-24. ISSN 2063-9309

67. HATCHER, L.; O'ROURKE, N. (2013): A step-by-step approach to using the SAS System for factor analysis and structural equation modelling. Second Edition. North Carolina, USA. SAS Institute, Inc. Last accessed on 16/10/2016 17:03 from: support.sas.com/publishing/pubcat/chaps/61314.pdf

68. HAMMER, M.; CHAMPY, J. (2001): Reengineering the corporation: A manifesto for business revolution. HarperCollins Publisher Inc., New York., pp. 9-33., ISBN: 0-06-662112-7

69. HANYECZ, L. (2006): A controlling rendszere. Saldo Kiadó, Budapest.

70. HENZLER, H. (1974): Der Januskopf muss weg! *Wirtschaftwoche.* 1974, No (38) pp. 60-63. ISSN: 0042-8582

71. HEY, T.; TANSLEY, S.; TOLLE, K. (2009): The fourth paradigm: Data-intensive Scientific Discovery, Microsoft Research, USA, pp. 15-22. ISBN: 978-0-9825442-0-4

72. HOFSTEDE, G.H. (1968): The game of Budget Control, Koninklikje Van Gorcum & Comp. N.V., Assen, The Netherlands, pp. 1-7 ISBN: 978-0-415-26510-2

73. HORVÁTH, P. (1986): Die Aufgaben des Marketing-Controllers, in: Eschenbach, R. (Hrsg.): Marketingcontrolling, Tagungsbericht Österreichischer Controllertag, Wien, pp. 7-29.

74. HORVÁTH, P. (1997): Controlling: a sikeres vezetés eszköze. Közgazdasági és Jogi Kiadó, Budapest.

75. HORVÁTH & PARTNER (2001): Controlling - Út egy hatékony controlling rendszerhez, KJK Kerszöv Kiadó Kft., Budapest, p.72-143. ISBN: 978-9-6322-4940-7

76. HORVÁTH, P. (2003): A Controller: a vezetés navigátora, Menedzsment Fórum Online, Last accessed on 02/10/2016 18:11 from http://www.mfor.hu/cikkek/9708.html

77. HORVÁTH & PARTNER (2004): Controlling értelmező szótár, Gyomai Kner Nyomda., Budapest, ISBN: 963-214-861-4

78. HUANG, X.; ZHANG, M.; HUI, M.K.; WYER, R.S. Jr. (2014): *Warmth and conformity: The effects of ambient temperature on product preferences and financial decisions.* Journal of Consumer Psychology. Vol(24) No(2), pp 241-250. ISSN: 1057-7408 Last downloaded: 23/11/2016 21:49 from www.bm.ust.hk/mark/staff/Bob/mkwyer-JCP-2014-24-241.pdf

79. ILDZSA, SZ. (2009): A Magyar Fejlesztési Bank controlling rendszere, mint versenyképességi tényező Egyetemi Diploma dolgozat, Szent István Egyetem, Gödöllő pp. 7-38.

80. JÓZSA, L.; PISKÓTI, I.; REKETTYE, G.; VERES, Z. (2005): Döntésorientált Marketing, KJK Kerszöv, Budapest, pp. 395-398. ISBN:978-9-6322-4784-7

81. JÓZSA, L. (2006): A marketingcontrolling döntéselőkészítő szerepe a vállalkozások marketingirányítási és –ellenőrzési feladatainál. *A Controller*, Vol (2), No (1), pp. 13-16. ISSN 2063-9309

82. KAISER, H.F. (1960): *The application of electronic computers to factor analysis*. Educational and Psychological measurement. Vol (20) No (1) pp.141-151 eISSN: 1552-3888 Last accessed on 17/10/2016 19:54 from: http://epm.sagepub.com/content/20/1/14

83. KALMÁR, P. (2010): Unsecured business loan NPV model for cut-off review, MSC Thesis, Cranfield University, United Kingdom

84. KALMÁR, P.; ILDZSA, SZ. (2012): A hitelintézeti marketing-kontrolling elméleti modellje. A jövő gazdasága, a jövő befektetése. Conference Proceedings, Gödöllő, 2012. Electronic edition (CD) ISBN: 978-963-269-326-2

85. KALMÁR P., ZÉMAN Z., LUKÁCS J. (2015) *Bankkontrolling marketing szemléletben – alkalmazott statisztika a kontrolling szolgálatában*. Hitelintézeti szemle, Budapest, Vol (14) No (4), pp 108-123. ISSN: 1588-6883

86. KANDIKÓ, J.; LEHEL, GY. (2012): Marketingszámtan – Tervezés és kontrolling a marketingben és az értékesítés-menedzsmentben (Főiskolai jegyzet), Edutus Főiskola, Tatabánya, pp. 4-20.; 67. Last accessed on 26/02/2017 13:53 from https://issuu.com/mutf/docs/marketingsz_mtan__2012_

87. KAPLAN, R.S.; NORTON, D.P. (1992): *The Balanced Scorecard: Measures that Drive Performance*. Harvard Business Review, USA, (January-February): pp. 71-79. ISSN: 0017-8012

88. KEINER, J. (1980): Marketing-Controlling (Controlling-Praxis), Toeche-Mittler, Darmstadt, pp. 7-23. ISBN: 978-3-8782-0042-0

89. KMENTA J. (1997): Elements of Econometrics - Second edition. New York, USA. Macmillan, pp.607-650. ISBN: 978-0472108862

90. KOTLER, P.; KELLER, K.L. (2012): Marketing Management – 14th edition. New York, USA. Prentice Hall, pp. 2-31.; 922-947. ISBN: 978-0-13-210292-6

91. KOVÁCS I. (1977): Az autokorreláció vizsgálata regressziós modellekben. *Statisztikai Szemle*. 552. évf. 6. sz. 603–621. old. Last downloaded: 23/09/2016 20:15 from http://www.ksh.hu/statszemle_archivum#author=Kov%C3%A1cs%20Ilona

92. KOVÁCS, L.; PÁL, ZS. (2012): A pénzügyi infrastruktúra fejlesztése és várható hatásai. Hitelintézeti szemle, Budapest, Vol (11) No (1), pp 47-59. ISSN: 1588-6883

93. KÖHLER, R. (1993): Beiträge zum Marketing-Management – 3 erweiterte Auflage, Schäffer-Poeschel Verlag, Stuttgart, Germany, pp. 78-152. ISBN: 978-3-7910-0682-6

94. KÖHLER, R. (2006): Marketingcontrolling – Konzepte und Methoden IN: TOMCZAK, S.; REINECKE, S. ed. Handbuch Marketing-Controlling 2nd edition, Gabler, Germany, pp. 41-55. ISBN: 978-3-8349-9335-9

95. KÖRMENDI, L.; TÓTH, A. (2006): A controlling elmélete és gyakorlata. Perfekt Gazdasági Tanácsadó, Oktató és Kiadó, Budapest, pp. 19-22., ISBN: 978-963-394-639-8

96. KUMAR, R. (2014): Strategies of Banks and Other Financial Institutions – Theories and Cases. Academic Press (imprint of Elseiver), San Diego., pp.3-29., ISBN: 978-0-12-416997-5

97. KÜPPER, H.U. (2007): Controlling und Operations Research - Der Beitrag quantitativer Theorie zur Selbstfindung und Akzeptanz einer praxisorientierten Disziplin. *Zeitschrift für Betriebswirtschaft*. Issue: 7/8, pp. 735-758. ISBN: 0044-2372

98. KAPLAN, R.S.; ATKINSON, M.Y.A. (2003): Magement Accounting: 4th edition. New Jersey: Prentice Hall. pp. 14-22. ISBN: 978-0273760160

99. KAPLAN, R.S.; JOHNSON, H.T. (1987): Relevance Lost: The Rise and Fall of Management Accounting. Boston: Harvard Business School Press. pp. 44-60. ISBN:0-87584-254-2

100. KERÉKGYÁRTÓ, GY.; BALOGH, L.I.; SUGÁR, A.; SZARVAS, B. (2009): Statisztikai módszerek és alkalmazásuk a gazdasági és társadalmi elemzésekben. Budapest, Aula Kiadó. pp. 340-375. ISBN:978-963-969-836-9

101. KEHL, D.; SIPOS, B. (2010): *Regressziós modellek becslése és tesztelése Excel-parancsfájl segítségével (szoftverismertetés).* Statisztikai Szemle. 88. évf. 7-8. sz. 833–855. old. Last downloaded: 23/09/2016 22:12 from http://www.ksh.hu/statszemle_archive/2010/2010_07-08/2010_07-08_833.pdf

102. KNOFCZYNKSI, G.T.; MUNDFROM, D. (2008) Sample Sizes When Using Multiple Linear Regression for Prediction. *Educational and Psychological Measurement*, June 2008 Vol (68), No (3), pp.:431-442. ISSN: 0013-1644

103. LANE, D.C. (2008): The emergence and use of diagramming in system dynamics: A critical Account. *Systems Research and Behavioural Science,* Vol (25) No (1), pp 3-23. ISSN: 1099-1743

104. LANEY, D. (2001): 3D Data Management: Controlling Data Volume, Velocity, and Variety. META group Inc., Accessed on 13/12/2013 from: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

105. LÁZÁR, L. (2002): Értékek és mértékek – a vállalati erőforrás-felhasználás leképezése és elemzése hazai üzleti szervezetekben. Doktori (PhD) értekezés. Budapesti Közgazdaságtudományi és Államigazgatási Egyetem, Gazdálkodástudományi PhD. Program, Budapest pp.: 41-55

106. LENTNER, CS. (2009): A válság néhány nemzetközi és pénzügypolitikai összefüggése, *Számvitel – Adó – Könyvvizsgálat,* Vol (51) No (9), pp. 372-376. ISSN: 1419-6956

107. LENTNER, CS. (2010): Some Historical and International Aspect of Hungarian Economic Crisis and Crisis Management. *Journal of Public Finance*, Vol (3/2010), No (3) pp. 581-606., ISSN: 1588-6883

108. LENTNER, CS. (2011): Pénzintézetek társadalmi felelősségvállalásának új dimenziói és a könyvvizsgálat szerepe, *Számvitel – Adó – Könyvvizsgálat*, Vol (53) No (7-8), pp. 348-350. ISSN: 1419-6956

109. LENTNER, CS. (2013): Bankmenedzsment - Bankszabályozás, pénzügyi fogyasztóvédelem, Nemzeti Közszolgálati és Tankönyv Kiadó Zrt, Budapest, pp. 312-344. ISBN: 978-9-630-855-9144

110. LENTNER, CS. (2015): A devizahitelezés nagy kézikönyve, Nemzeti Közszolgálati és Tankönyv Kiadó Zrt, Budapest, pp. 27-33. ISBN: 978-6-155-344-626

111. LENTNER, CS.; KOLOZSI, P.; TÓTH, G. (2010): A magyar válságkezelés sajátosságai és ellentmondásai. In: MAJOROS, P. (editor): EU Working Papers, Vol (1), No (1), Budapesti Gazdasági Főiskola Külkereskedelmi Főiskolai Kar Könyvtára, Budapest, pp. 1-18. ISSN: 1418-6241

112. LOWE, E.A.; TINKER, A.M. (1977): New directions for managment accounting. *The International Journal of Management Science*, Vol (5), No (2), pp. 173-183. ISSN: 0305-0483

113. MABBERLEY, J. (1999): A pénzügyi intézmények controllingja, Panem Kft., Budapest, pp. 13-23 ISBN: 963-545-191-1

114. MACINTOSH, N.B. (1981): A contextual model of information systems. *Accounting, Organization and Society*, Vol (6), No (1), pp. 39-59. ISSN: 0361-3682

115. MAHER, M. W., (2000): Management Accounting Education at the Millennium. *Issues in Accounting Education*. Vol (15), No (2), pp. 335-346. ISSN: 1558-7983

116. MANN, R.; MAYER E. (1993): Controlling kezdők számára. Saldo Rt., Budapest.

117. MCEWEN, A.; CASSIMALLY, H. (2014): Designing the Internet of Things, John Wiley and Sons Ltd. West Sussex, UK, pp. 7-20., ISBN: 978-1-118-43062-0

118. MEFFERT, H.; BURMANN, C.; KIRCHGEORG, M. (2015): Marketing: Grundlagen marktorientierter Unternehmensführung: Konzepte – Instrumente – Praxisbeispiel, Gabler Verlag, Wiesbaded, pp. 43-76. ISBN:978-3-658-02343-0

119. MERCHANT, K.A.; VAN DER STEDE, W.A. (2011): Management Control Systems: Performance Measurement, Evaluation and Incentives – Third Edition, Financial Times / Prentice Hall, Harlow, England, pp. 3-24; 218-268., ISBN: 978-0-2737-3761-2

120. MORGAN, G. (2006): Images of organization. Sage Publications Inc. California, USA. pp. 115-148. ISBN: 1-4129-3979-8

121. NEIMARK, M.; TINKER, T. (1986): The social construction of managment control system. *Accounting, Organization and Society*, Vol (11), No (4/5), pp. 369-395. ISSN: 0361-3682

122. NETER, J.; KUTNER, M.; NACHTSHEIM, C.; WASSERMAN, W. (1996): Applied Linear Statistical Models 4th Edition. McGraw-Hill/Irwin., USA, pp. 428-490. ISBN: 978-0256117363

123. NGAI, L.R.; TENREYRO, S. (2014): Hot and Cold Seasons in the Housing Market. *American Economic Review*, Vol (104) No (12), pp. 3991-4026. ISSN: 0002-8282

124. NORNADIAH, M. R.; YAP, B. W. (2011): Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modelling and Analytics*. Vol (2), No (1), pp. 21-33. eISSN: 14770342

125. OECD (2016): Business confidence index (BCI) Accessed on 26/11/2016 from: https://data.oecd.org/leadind/business-confidence-index-bci.htm

126. OTLEY, D. (1994): Management control in contemporary organizations: towards a wider framework. *Management Accounting Research*, Vol (5), pp. 289-299. ISSN: 1044-5005

127. OTLEY, D. (1999): Performance Management: a framework for management control systems research. *Management Accounting Research*, Vol (10), pp. 363-382. ISSN: 1044-5005

128. PARDOE, I.; SIMON, L.; YOUNG, D. (2016): STAT 501 – Regression Methods. Penn State Eberly College of Science. Online course meterials website. Last accessed on 06/10/2016 from: https://onlinecourses.science.psu.edu/stat501/node/2

129. PARKER, H.R. (1969): Management Accounting: an historical perspective. New York, A.M. Kelley. pp. 21-69. ISBN: 978-0678-070-017

130. PEDHAZUR, E. J.; SCHMELKIN, L. P. (1991). Measurement, design, and analysis: An integrated approach. New York, Taylor & Francis Group pp. 318-341. ISBN: 978-0-8058-1063-9

131. PICARD, R.R.; COOK, R.D. (1984) Cross-Validation of Regression Models. *Journal of the American Statistical Association* Vol (79), No (387), pp. 575-83. ISSN:1537-274X Downloaded from: http://www.jstor.org/stable/2288403 Accessed on 23/10/2016

132. PIERCY, N.F. (1987): The Marketing Budgeting Process: Marketing Management Implications. *Journal of Marketing*. Vol (51) No (4), pp. 45-49. ISSN: 0022-2429

133. REICHMANN, T. (2001): Controlling mit Kennzahlen und Managementberichten (6th edition), Verlag Vahlen, Munich, pp. 53-77., ISBN:978-3-8006-2531-4

134. REINECKE, S. (2006): Effektivität und Effizienz einer marktorientierten Unternehmensführung IN: TOMCZAK, S.; REINECKE, S. ed. Handbuch Marketing-Controlling 2nd edition, Gabler, Germany, pp. 5-40. ISBN: 978-3-8349-9335-9

135. REINECKE, S.; FUCHS, D. (2006): Marketingbudgetierung IN: TOMCZAK, S.; REINECKE, S. ed. Handbuch Marketing-Controlling 2nd edition, Gabler, Germany, pp. 795-820. ISBN: 978-38349-93359

136. REINECKE, S.; JANZ, S. (2007): Marketing-Controlling (Scherstellen von Marketingeffektivitat und -effizenz), Kohlhammer, Stuttgart, Germany, pp. 395-419. ISBN: 978-3-17-018404-6

137. RIEG, R. (2015): Planung und Budgetierung (2nd edition), Gabler Verlag, Wiessbaden, pp. 193-218. ISBN: 978-3-8349-4628-7

138. RIKHARDSSON, P.; CHRISTENSEN, L.; ROHDE, C.; BATT, C.E. (2016): Financial crisis and changes in management controls in banks, Chartered Institute of Management Accountants, London, pp. 1-16. ISSN: 1744-7038

139. ROSANAS, J.M. (1994): Coming to a new awareness of oranizational culture. *Sloan Management Review*. Vol (25), No (2), pp. 3-16. ISSN 15329194

140. ROTHER, M. (2009): Toyota Kata: managing people for improvement, adaptiveness, and superior results, McGraw-Hill, USA, pp.133-135. ISBN: 978-0-07-163523-3

141. SÁGI, J. (2007): Banktan, Saldo Zrt., Budapest, pp. 63-171 ISBN: 978-9-6363-8235-3

142. SAJTOS, L.; MITEV, A. (2007): SPSS kutatatási és adatelemzési kézikönyv. Budapest. Alinea Kiadó. pp.245-282 ISBN: 978-963-9659-08-7

143. SCHMALENBACH, E. (1963): Kostenrechnung und Preispolitik. Westdeutscher Verlag, Köln/Opladen

144. SCHWALBE, H. (1990): Gyakorlati marketing kis- és középvállalkozások számára. Közgazdasági és Jogi Kiadó, Budapest.

145. SHAPIRO, S.S.; WILK, M.B. (1965): An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, Vol(52), No(3-4), pp. 591-611. ISSN:0006-3444

146. SNEDECOR, G.W.; COCHRAN, W.G. (1989): Statistical Methods, Eighth Edition, Iowa State University Press, pp.217-235. ISBN:0-8138-1561-4

147. SPREMANN, K. (1992): Grundlagen - Informationssysteme – Anwendungen. Gabler, Wiesbaden. ISBN: 978-340913-1384

148. STIRES, C. (2013): Big Data Ecosystem – From Data to decision, IDC's Asia/Pacific Big Data Conference 2013, IDC, Melbourne, Australia 03-Sep-2013, Last accessed on 05/03/2016 20:05 from: http://infographics.idc.asia/bigdata/Big_Data_ecosystem.asp

149. STRAUSS, R.E. (2008): Marketing planning by design – Systematic Planning for Successful Marketing Strategy, John Wiley & Sons Ltd. West Sussex, England, pp.71-74. ISBN: 978-0-470-72167-4

150. SUHR, D.D. (2005): Principal Component Analysis vs. Exploratory Factor Analysis in SAS Institute Inc. 2005. Proceedings of the Thirtieth Annual SAS® Users Group International Conference. Cary, NC: SAS Institute Inc. Paper: 203-30. Last accessed on 11/10/2016 08:12 from: http://www2.sas.com/proceedings/sugi30/203-30.pdf

151. SZÉKELYI, M.; BARNA, I. (2002): Túlélőkészlet az SPSS-hez – Többváltozós elemzési technikákról társadalomkutatók számára. Budapest, Typotex Kiadó pp. 17-300. ISBN 963-9326-42-9

152. SZÓKA, K. (2007): A pénzügyi-számviteli tervezés és a controlling összefüggései és gyakorlata (Különös tekintettel a kis- és középvállalkozásokra). Doktori (PhD) értekezés. Nyugat-Magyarországi Egyetem, Sopron. pp. 84.-118.

153. SZŰCS, I. (2004): Alkalmazott Statisztika, Budapest, Agroinform Kiadó és Nyomda Kft. pp. 151-179., 275-550. ISBN: 963-502-761

154. SZŰCS, R. (2014): A marketing controlling, *Számviteli Tanácsadó* Vol (6), No (7/8), ISSN: 2060-4076 Last accessed from: http://ado.hu/szaklap/szamviteli-tanacsado/2014-7-8-2/a-marketing-kontrolling on 01/02/2016 13:52

155. SWARTZ, T.A.; IACOBUCCI, D. (2000): Handbook of services marketing and management, London, Sage Publications, Inc. pp. 150-158., ISBN: 0-7619-1611-3

156. TAYLOR, F.W. (1911): The Principles of Scientific Management – Reprint (2006), Cosimo Inc, New York, USA. ISBN: 1-59605-889-7

157. TANNENBAUM, A. S. (1966): Organizational theory and organizational practice. *Management International Review*. Vol (5), pp. 121-135. ISSN 0938-8249

158. THEIL, H. (1961): Economic Forecasts and Policy (Contributions to Economic Analysis) – 2nd edition. North-Holland Pub. Co. pp. 124-209. ISBN: 978-0720431285

159. TÓTH, A. (1998): A kontrolling szerepe a kórházak irányításában. *Kórház*, Vol (5), No (4) pp. ISSN: 0230-3868

160. TÓTH, A., ZÉMAN, Z. (2003): A konszerncontrolling gyakorlatimegvalósításának lehetőségei. *Pénzügyi Szemle*. 48. évf. 8.sz. / 2003 pp. 775-790. ISSN 2064-8278

161. TÓTH, A., ZÉMAN, Z. (2006): Szervezeti Controlling – Egyetemi jegyzet, Szent István Egyetem, Gyöngyös, pp. 1-140.

162. TÓTH, G. (2013): A monetáris politika új lehetőségei. Gazdaság és Jog. Vol (21), No (5), pp. 10-13. ISSN 1217-2464

163. TÓTH, G., JUHÁSZ, G. (2014): Jegybankok a világ megmentői. *Gazdasági Élet és Társadalom*. 6. évf. I-II.sz. pp. 99-109. ISSN 2060-7466

164. TURNER, V.; GANTZ, J.F.; REINSEL, D.; MINTON, S (2014): The digital universe of opportunities: rich data and the increasing value of the internet of things. IDC White paper. Last accessed on 05/03/2016 from http://idcdocserv.com/1678

165. VAN IWAARDEN, J.; WILLIAMS, R.; DALE, B. (2006): A management control perspective of quality management. *International Journal of Quality & Reliability Management*, Vol (23), No (1), pp. 102–112. ISSN: 0265-671X Last accessed on 18/01/2017 doi: 10.1108/02656710610637578.

166. VÉRY, Z. (2009): Funkcionális controlling, Budapest, Raabe Kiadó, pp. 3-27., 77-98.

167. WARNER, M.R. (1998): Spectral Analysis of Time-Series Data. New York, USA, The Guilford Press. pp. 78-97., ISBN:978-1572303386

168. WEBER, J; SHAFFER, U. (2006): Marketingcontrolling: Sicherstellung der Rationalität in einer marktorientierten Unternehmensführung IN: TOMCZAK, S.; REINECKE, S. ed. Handbuch Marketing-Controlling 2nd edition, Gabler, Germany, pp. 32-39. ISBN: 978-3-8349-9335-9

169. WELCH, M. (2011): The evolution of the employee engagement concept: communication implications. *Corporate Communications: An International Journal*, Vol (16) Iss: 4, pp.328-346. ISSN: 1356-3289

170. WOOLDRIDGE, J. M. (2002): Econometric Analysis of Cross Section and Panel Data. Massachusetts, USA, MIT Press. pp.279-291., ISBN:978-0-262-23219-7

171. YEARWORTH, M. (2014): A Brief Introduction to System Dynamics Modelling pp.15. University of Bristol. Available online: http://www.groundedsystems.com/wp-content/uploads/2015/02/SD-Introduction-MY-241014.pdf Last downloaded: 18/09/2016 18:46

172. YONG, A.G.; PEARCE, S. (2013): A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. Tutorials in Quantitative Methods for Psychology. Vol (9) No (2), pp. 79-94. ISSN: 1913-4126

173. ZÉMAN, Z. (1998): A Controlling szerepe a magyar vállalkozásokban, Doktori (PhD) értekezés. Szent István Egyetem, Gödöllő. pp. 8.-69.

174. ZÉMAN, Z.; BÉHM, I. (2016): A pénzügyi menedzsment controll elemzési eszköztára, Akadémiai Kiadó, Budapest, pp. 13-36. ISBN: 978-9-630-597-746

175. ZÉMAN, Z.; GACSI, R.; LUKÁCS, J.; HAJÓS, L. (2013): Management control system in banks. *BIATEC – banking journal*, Vol (21) No (6), pp 14-17. ISSN: 1335-0900

Other References

176. European Commission (2010): Communication from the Commission: Europe 2020 – A European strategy for smart, sustainable and inclusive growth. COM 2020

177. European Commission (2013): Digital Agenda for Europe, Session Reports, ICT for Industrial Leadership: Innovating by exploiting big and open data and digital content

178. European Commission (2014): Towards a thriving data-driven economy, Communication from the commission to the European Parliament, the council, the European economic and social Committee and the committee of the regions, Brussels

179. KOVÁCS, L. (2012): Hitelintézeti fogalomtár, Budapest, Magyar Bankszövetség, pp. 7-182. ISBN 978-963-89653-0-1

180. OECD (2013): Exploring Data-Driven Innovation as a New Source of Growth – mapping the policy issues raised by "Big Data." Rep. from OECD.

A2. Table of Figures

A3. List of Tables

A4. List of Acronyms

Abbrev.	Name
ABC	Activity Based Costing
ACF	Autocorrelation Function
AER	Annual Equivalent Rate
ANOVA	Analysis of Variance
API	Application Programing Interface
APR	Annual Percentage Rate
AR	Autoregression
ARIMA	Autoregressive Integrated Moving Avg.
B	Byte
BBC	British Broadcasting Company
BCI	Business Confidence Index
BPR	Business Process Reengineering
C	C programming language
CA	Current Account
CACI	The name of a market research agency
CC	Credit Card
CCI	Consumer Confidence Index
CET1	Tier 1 Common Capital Ratio
CFA	Common Factor Analysis
CIB	Corporate & Institutional Banking
CIMA	Chartered Institute of Management Accountants
CPB	Commercial & Private Banking
CRM	Customer Relationship Management
DF	Degree of Freedom
DFBETA	The measure of how much impact each observation has on a predictor
DW	Durbin-Watson
EFA	Exploratory Factor Analysis
EMU	European Monetary Union
GB	Giga Byte
GBP	Great British Pound
GCSE	General Certificate of Secondary Education
GDI	Gross Disposable Income
GDP	Gross Domestic Product
GFS	Global Financial Services
GVA	Gross Value Added
HSBC	Hongkong and Shanghai Banking Corp.
IAAS	International Auditing and Assurance Standards
IAS	International Accounting Standards
ISA	Individual Savings Account
IT	Information Technology
KMO	Kaiser-Meyer-Olkin
KSH	Központi Statisztikai Hivatal

Abbrev.	Name
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MB	Mega Byte
METOFFICE	The national weather service
MFI	Monetary Financial Institutions
MINEIGEN	Minimum Eigen Value
MNB	Magyar Nemzeti Bank
MSA	Measure of Sampling Adequacy
MSPE	Mean Squared Prediction Error
NPS	Net Promoter Score
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
ONS	Office for National Statistics
PACF	Partial Autocorrelation Function
PBB	Personal & Business Banking
PCA	Principal Component Analysis
PDCA	Plan, Do, Check, Act
PI	Prediction Interval
PLC	Public Limited Company
PPI	Payment Protection Insurance
RAM	Random Access Memory
RBS	Royal Bank of Scotland
RMS	Root Mean Square
RMSE	Root Mean Square Error
ROI	Return on Investment
SAS	Statistical Analysis System
SBC	Schwarz Bayesian Criterion
SQL	Structured Query Language
SSE	Error Sum of Squares
SSR	Sum of Squares
SSRN	Social Science Research Network
SSTO	Total Sum of Squares
SVR	Standard Variable Rate
SWOT	Strenght, Weakness, Opportunity, Threat
TD	Telephone Distribution
TV	Television
ULS	Unweighted Least Squares
UPL	Unsecured Personal Loan
USA	United States of America
USD	US Dollar
VIF	Variance Inflation Factor
ZB	Zetta Byte

10.14751/SZIE.2017.056

A5. Detailed List of Initial Predictor Variables and their Methodologies

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Group 1. General Operations	Infrastructure	Branches	Bank A Branches	INFRASTRUCTURE_BANK_BRANCHES	Internal - Management Information System	Calculate the end of week stock positions
			Competitor Branches	INFRASTRUCTURE_COMPETITOR_BRANCHES	External - Market Research Agency	Weekly phasing from monthly data
			Branch Share	INFRASTRUCTURE_BRANCH_SHARE	-	(Bank A branches)/(Bank A + Copetitor branches)
		Staff	All	INFRASTRUCTURE_STAFF_ALL	Internal - HR Administration System	Calculate the end of week stock positions
			Branch	INFRASTRUCTURE_STAFF_BR	Internal - HR Administration System	Calculate the end of week stock positions
			TD	INFRASTRUCTURE_STAFF_TD	Internal - HR Administration System	Calculate the end of week stock positions
	Operations and Service Quality	Branch	Branch Completed Appointments	OPERATIONS_BR_COMPL_APPOINTMENTS	Internal - Management Information System	Calculate weekly total appointments completed
		TD	Completed Appointments	OPERATIONS_TD_APPOINTMENTS	Internal - Management Information System	Calculate weekly total appointments completed
		Internet	Website Visits	OPERATIONS_INT_VISITS	Internal - Adobe Analytics	Calculate weekly total visits (1 visit = 1 session) to the website
		Customer Satisfaction	Bank A Internal Cust. Satisfaction	OPERATIONS_CUST_SAT_INTERNAL	Internal - Management Information System	(Total Satisfied - Total Dissatisfied)/(Total number of complete surveys)
			Bank A Excernal Cust. Satisfaction	OPERATIONS_CUST_SAT_MARKET	External - Market Research Agency	(Total Satisfied - Total Dissatisfied)/(Total number of complete surveys)
			Service level competittieveness	OPERATIONS_CUST_SAT_COMPETITVE	-	Ranking on weekly Customer Satisfaction lead table
	Risk Appetite	Accept rate	Current Account	RISK_CA_ACCEPT_RATE	Internal - Risk Information System	Accepted Weekly CC and CA Applications / All Weekly Applications

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Group 2. Product and Market Share	Product Openings by channel	Current Account Openings (split by customer type)	Banking Liability (£m)	PRODUCTS_CA_LIABILITY	Internal - Management Information System	Calculate the end of week stock positions
			Flagship CA Liability (£m)	PRODUCTS_CA_FLAGSHIP_LIABILITY	Internal - Management Information System	Calculate the end of week stock positions
			Current Account Openings New - Total	PRODUCTS_CA_TOTAL_NEW_BY_NTG	Internal - Management Information System	Calculate weekly total openings by new to banking group main customer
				PRODUCTS_CA_TOTAL_NEW_BY_NTB	Internal - Management Information System	Calculate weekly total openings by new to banking main customer
				PRODUCTS_CA_TOTAL_NEW_BY_EXIST	Internal - Management Information System	Calculate weekly total openings by existing banking main customer
				PRODUCTS_CA_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
			Current Account Openings New - Branch	PRODUCTS_CA_BRANCH_NEW_BY_NTG	Internal - Management Information System	Calculate weekly total openings by new to banking group main customer
				PRODUCTS_CA_BRANCH_NEW_BY_NTB	Internal - Management Information System	Calculate weekly total openings by new to banking main customer
				PRODUCTS_CA_BRANCH_NEW_BY_EXIST	Internal - Management Information System	Calculate weekly total openings by existing banking main customer
				PRODUCTS_CA_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
			Current Account Openings New - TD	PRODUCTS_CA_TD_NEW_BY_NTG	Internal - Management Information System	Calculate weekly total openings by new to banking group main customer
				PRODUCTS_CA_TD_NEW_BY_NTB	Internal - Management Information System	Calculate weekly total openings by new to banking main customer
				PRODUCTS_CA_TD_NEW_BY_EXIST	Internal - Management Information System	Calculate weekly total openings by existing banking main customer
				PRODUCTS_CA_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
			Current Account New Openings - Internet	PRODUCTS_CA_INTERNET_NEW_BY_NTG	Internal - Management Information System	Calculate weekly total openings by new to banking group main customer
				PRODUCTS_CA_INTERNET_NEW_BY_NTB	Internal - Management Information System	Calculate weekly total openings by new to banking main customer
				PRODUCTS_CA_INTERNET_NEWS_BY_EXIST	Internal - Management Information System	Calculate weekly total openings by existing banking main customer
				PRODUCTS_CA_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
			Current Account Upgrade - Total	PRODUCTS_CA_UPGRADES_BY_BRANCH	Internal - Management Information System	Calculate weekly total banking upgrades to flagship product for branch
				PRODUCTS_CA_UPGRADES_BY_TD	Internal - Management Information System	Calculate weekly total banking upgrades to flagship product for TD
				PRODUCTS_CA_UPGRADES_BY_INTERNET	Internal - Management Information System	Calculate weekly total banking upgrades to flagship product for internet
				PRODUCTS_CA_UPGRADES_BY_ALL	Internal - Management Information System	Calculate weekly total banking upgrades to flagship product
		Credit Card	CC Branch	PRODUCTS_CC_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			CC Internet	PRODUCTS_CC_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for internet
			CC TD	PRODUCTS_CC_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			CC Total	PRODUCTS_CC_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
		General Insurance	GI Branch	PRODUCTS_GI_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			GI Internet	PRODUCTS_GI_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for internet
			GI TD	PRODUCTS_GI_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			GI Total	PRODUCTS_GI_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
		Investment	INV Branch	PRODUCTS_INV_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			INV Total	PRODUCTS_INV_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
		Mortgage	MR Branch	PRODUCTS_MR_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			MR TD	PRODUCTS_MR_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			MR Total	PRODUCTS_MR_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Continue: Group 2. Product and Market Share	Continue: Product Openings by channel	Savings - Instant Access and ISA	SAV Branch	PRODUCTS_SAV_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			SAV Internet	PRODUCTS_SAV_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for internet
			SAV TD	PRODUCTS_SAV_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			SAV Total	PRODUCTS_SAV_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
		Savings - Fixed Term	SAVFT Branch	PRODUCTS_SAVFT_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			SAVFT Internet	PRODUCTS_SAVFT_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for internet
			SAVFT TD	PRODUCTS_SAVFT_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			SAVFT Total	PRODUCTS_SAVFT_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
		UPL	UPL Branch	PRODUCTS_UPL_BRANCH_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for branch
			UPL Internet	PRODUCTS_UPL_INTERNET_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for internet
			UPL TD	PRODUCTS_UPL_TD_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings for TD
			UPL Total	PRODUCTS_UPL_TOTAL_NEW_BY_ALL	Internal - Management Information System	Calculate weekly total openings
	Market Share By Product And Channel	CA Market Share	Bank A Stock	MARKET_SHARE_CA_BANK_STOCK_SHARE	External - Market Research Agency	End of week volume stock of Bank A CAs (phasing to weekly)
			Market Stock	MARKET_SHARE_CA_MARKET_STOCK_SHARE	External - Market Research Agency	End of week volume stock of Total Market CAs (phasing to weekly)
			Stock Perc	MARKET_SHARE_CA_STOCK_SHARE_PERC	External - Market Research Agency	Bank A stock / Market Stock
			Bank A New Business	MARKET_SHARE_CA_BANK_NB_SHARE	External - Market Research Agency	End of week Year to Date new business volume of Bank A CAs (phasing)
			Market New Business	MARKET_SHARE_CA_MARKET_NB_SHARE	External - Market Research Agency	End of week Year to Date new business volume of Market CAs (phasing)
			Total Share	MARKET_SHARE_CA_SHARE_TOTAL	External - Market Research Agency	Bank A New Business (NB) Volumes / Market New Business (NB) Volumes
			Branch	MARKET_SHARE_CA_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes
			TD	MARKET_SHARE_CA_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes
			Online	MARKET_SHARE_CA_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes
		Savings - Fixed Term Market Share	Total Share	MARKET_SHARE_SAV_FT_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for Savings FT
			Branch	MARKET_SHARE_SAV_FT_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for Savings FT
			TD	MARKET_SHARE_SAV_FT_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for Savings FT
			Online	MARKET_SHARE_SAV_FT_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes for Savings FT
		Savings Market Share	Total Share	MARKET_SHARE_SAV_OTH_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for Savings liability
			Branch	MARKET_SHARE_SAV_OTH_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for Savings liab.
			TD	MARKET_SHARE_SAV_OTH_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for Savings liability
			Online	MARKET_SHARE_SAV_OTH_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes for Savings liab.
		General Insurance Market Share	Total Share	MARKET_SHARE_GI_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for Insurance vol.
			Branch	MARKET_SHARE_GI_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for Insurance
			TD	MARKET_SHARE_GI_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for Insurance volume
			Online	MARKET_SHARE_GI_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes for Insurance
		Credit card Market Share	Total Share	MARKET_SHARE_CC_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for CC volume
			Branch	MARKET_SHARE_CC_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for CC volume
			TD	MARKET_SHARE_CC_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for CC volume
			Online	MARKET_SHARE_CC_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes for CC volume
		UPL Market Share	Total Share	MARKET_SHARE_UPL_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for UPL assett
			Branch	MARKET_SHARE_UPL_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for UPL assett
			TD	MARKET_SHARE_UPL_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for UPL assett
			Online	MARKET_SHARE_UPL_SHARE_ONLINE	External - Market Research Agency	Bank A Online NB Volumes / Market Online NB Volumes for UPL assett
		Mortgage Market Share	Total Share	MARKET_SHARE_MR_SHARE_TOTAL	External - Market Research Agency	Bank A Total NB Volumes / Market Total NB Volumes for Mortg. assett
			Branch	MARKET_SHARE_MR_SHARE_BRANCH	External - Market Research Agency	Bank A Branch NB Volumes / Market Branch NB Volumes for M. assett
			TD	MARKET_SHARE_MR_SHARE_TD	External - Market Research Agency	Bank A TD NB Volumes / Market TD NB Volumes for Mortgage assett
	Market Competitiveness	Product competitiveness	Savings - Cash ISA position	MARKET_COMP_SAV_ISA_1_Yr_FR	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
				MARKET_COMP_SAV_ISA_2_Yr_FR	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
				MARKET_COMP_SAV_ISA_Instant	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
			Savings - Fixed Rate Bond position	MARKET_COMP_SAV_BOND_1_Yr	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
				MARKET_COMP_SAV_BOND_2_Yr	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
				MARKET_COMP_SAV_BOND_3_Yr	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)

10.14751/SZIE.2017.056

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Continue: Group 2. Product and Market Share	Continue: Market Competitiveness	Continue: Product competitiveness	Savings - Instant Access position	MARKET_COMP_SAV_INSTANT_Internet	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
				MARKET_COMP_SAV_INSTANT_Branch	External - Market Research Agency	Rank based on rate (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)
			Current Account Switchers Position	MARKET_COMP_CA_SWITCHERS	External - Market Research Agency	Net weekly total of switchers (Switched to - Switched out from Bank A)
		Press	Expert Mentions	MARKET_COMP_EXPERT_MENTIONS	External - Market Research Agency	Weekly Number of mentiones by Martin Lewis
			Google Searches Index	MARKET_COMP_GOOGLE_SEARCH	External - Google Inc.	Weekly Number of searches for the Bank A brand on www.google.co.uk
			Best Buy Mentiones	MARKET_COMP_BEST_BUY_MENTION	External - Market Research Agency	Weekly Number of mentions on the Best Buy Tables*
			Best Buy Shares	MARKET_COMP_BEST_BUY_SHARE	External - Market Research Agency	Weekly Number of Bank A mentions / Weekly Number of all mentions
			Best Buy Ranks	MARKET_COMP_BEST_BUY_RANK	External - Market Research Agency	Rank based on mentions (highest=4; 2nd=3; 3rd=2; 4th or more=5, none=0)

*For more information on best buy tables please see http://moneyfacts.co.uk/

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Group 3: Marketing Activity	Media Spend By Product Type	Total	Total	SPEND_MARKET_TOTAL	External - Market Research Agency	Weekly Total Market Media spend by agency method**
			Bank A	SPEND_MARKET_BANK	External - Market Research Agency	Weekly Total Bank A Media spend by agency method**
			Share	SPEND_MARKET_SHARE	External - Market Research Agency	Total Bank A / Total Market media spend by agency method**
		Brand Building/services	Total	SPEND_PRODUCT_BRAND_TOTAL	External - Market Research Agency	Weekly Market Brand spend by agency method**
			Bank A	SPEND_PRODUCT_BRAND_BANK	External - Market Research Agency	Weekly Bank A Brand spend by agency method**
			Share	SPEND_PRODUCT_BRAND_SHARE	External - Market Research Agency	Bank A / Total Market Brand spend by agency method**
		Business Banking Services	Total	SPEND_PRODUCT_BUSBA_TOTAL	External - Market Research Agency	Weekly Market Media spend on Business Banking by agency method**
			Bank A	SPEND_PRODUCT_BUSBA_BANK	External - Market Research Agency	Weekly Bank A Media spend on Business Banking by agency method**
			Share	SPEND_PRODUCT_BUSBA_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Bus. Ba. by agency method**
		General Insurance	Total	SPEND_PRODUCT_GI_TOTAL	External - Market Research Agency	Weekly Market Media spend on General Insurance by agency method**
			Bank A	SPEND_PRODUCT_GI_BANK	External - Market Research Agency	WeeklyBank A Media spend on General Insurance by agency method**
			Share	SPEND_PRODUCT_GI_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on GI by agency method**
		Investments	Total	SPEND_PRODUCT_INV_TOTAL	External - Market Research Agency	Weekly Market Media spend on Investment by agency method**
			Bank A	SPEND_PRODUCT_INV_BANK	External - Market Research Agency	Weekly Bank A Media spend on Investment by agency method**
			Share	SPEND_PRODUCT_INV_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Investment by agency method**
		Current Account	Total	SPEND_PRODUCT_CA_TOTAL	External - Market Research Agency	Weekly Market Media spend on Current Account by agency method**
			Bank A	SPEND_PRODUCT_CA_BANK	External - Market Research Agency	Weekly Bank A Media spend on Current Account by agency method**
			Share	SPEND_PRODUCT_CA_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on CA by agency method**
		Mortgages	Total	SPEND_PRODUCT_MORTG_TOTAL	External - Market Research Agency	Weekly Market Media spend on Mortgages by agency method**
			Bank A	SPEND_PRODUCT_MORTG_BANK	External - Market Research Agency	Weekly Bank A Media spend on Mortgages by agency method**
			Share	SPEND_PRODUCT_MORTG_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Mortgages by agency method**
		Unsecured Personal Loans	Total	SPEND_PRODUCT_UPL_TOTAL	External - Market Research Agency	Weekly Market Media spend on Unsecured Loan by agency method**
			Bank A	SPEND_PRODUCT_UPL_BANK	External - Market Research Agency	Weekly Bank A Media spend on Unsecured Loan by agency method**
			Share	SPEND_PRODUCT_UPL_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on UPL by agency method**
		Credit Cards	Total	SPEND_PRODUCT_CC_TOTAL	External - Market Research Agency	Weekly Market Media spend on Credit Card by agency method**
			Bank A	SPEND_PRODUCT_CC_BANK	External - Market Research Agency	Weekly Bank A Media spend on Credit Card by agency method**
			Share	SPEND_PRODUCT_CC_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Credit Card by agency method**
		Savings	Total	SPEND_PRODUCT_SAV_TOTAL	External - Market Research Agency	Weekly Market Media spend on Savings by agency method**
			Bank A	SPEND_PRODUCT_SAV_BANK	External - Market Research Agency	Weekly Bank A Media spend on Savings by agency method**
			Share	SPEND_PRODUCT_SAV_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Savings by agency method**
	Media Spend By Media Type	Cinema	Total	SPEND_MEDIA_CINEMA_TOTAL	External - Market Research Agency	Weekly Market Media spend on Cinema by agency method**
			Bank A	SPEND_MEDIA_CINEMA_BANK	External - Market Research Agency	Weekly Bank A Media spend on Cinema by agency method**
			Share	SPEND_MEDIA_CINEMA_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Cinema by agency method**
		Digital	Total	SPEND_MEDIA_DIGITAL_TOTAL	External - Market Research Agency	Weekly Market Media spend on Digital by agency method**
			Bank A	SPEND_MEDIA_DIGITAL_BANK	External - Market Research Agency	Weekly Bank A Media spend on Digital by agency method**
			Share	SPEND_MEDIA_DIGITAL_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Digital by agency method**

**The data provider Market Research Agency collect data 24 hours a day for 7 days a week and based on an extensive database of cost associated to certain marketing activity they collect, categorise and estimate money spend on each media activity by product, media, brand, etc. For example, for press activity all key newspapers are reviewed and based on the day of the advertisement (which influences the amount of copies sold), the page and size a cost is estimated for that day for the advertiser institution split by product. All these spends then collected and aggregated up by product/campaign type and media type.

Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Continue: Group 3: Marketing Activity	Continue: Media Spend By Media Type	Direct Mail	Total	SPEND_MEDIA_DM_TOTAL	External - Market Research Agency	Weekly Market Media spend on Direct Mail by agency method**
			Bank A	SPEND_MEDIA_DM_BANK	External - Market Research Agency	Weekly Bank A Media spend on Direct Mail by agency method**
			Share	SPEND_MEDIA_DM_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Direct Mail by agency method**
		Outdoor	Total	SPEND_MEDIA_OUTDOOR_TOTAL	External - Market Research Agency	Weekly Market Media spend on Outdoor by agency method**
			Bank A	SPEND_MEDIA_OUTDOOR_BANK	External - Market Research Agency	Weekly Bank A Media spend on Outdoor by agency method**
			Share	SPEND_MEDIA_OUTDOOR_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Outdoor by agency method**
		Press	Total	SPEND_MEDIA_PRESS_TOTAL	External - Market Research Agency	Weekly Market Media spend on Press by agency method**
			Bank A	SPEND_MEDIA_PRESS_BANK	External - Market Research Agency	Weekly Bank A Media spend on Press by agency method**
			Share	SPEND_MEDIA_PRESS_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Press by agency method**
		Radio	Total	SPEND_MEDIA_RADIO_TOTAL	External - Market Research Agency	Weekly Market Media spend on Radio ads. by agency method**
			Bank A	SPEND_MEDIA_RADIO_BANK	External - Market Research Agency	Weekly Bank A Media spend on Radio ads. by agency method**
			Share	SPEND_MEDIA_RADIO_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on Radio ads. by agency method**
		TV	Total	SPEND_MEDIA_TV_TOTAL	External - Market Research Agency	Weekly Market Media spend on TV advs. by agency method**
			Bank A	SPEND_MEDIA_TV_BANK	External - Market Research Agency	Weekly Bank A Media spend on TV advs. by agency method**
			Share	SPEND_MEDIA_TV_SHARE	External - Market Research Agency	Bank A / Total Market Media spend on TV advs. by agency method**
	Marketing Activity	Campaign Activity	Sponsorship Event	SPONSORHIP_CUMMULATIVE	Internal - Management Information System	Weekly total number of distinctive sponsorship events of Bank A
			Direct Marketing	DIRECT_MARKETING_SIZE	Internal - Management Information System	Weekly total number of customers e-mailed and/or direct-mailed
			TV Campaign	TV_CAMPAIGN_START_BINARY	Internal - Management Information System	Weekly binary flag of active TV campaign (at least 1 day = 1, none = 0)
	Brand Power	Brand Indexes	Unaided awareness - all mentions	BRAND_AWARENESS_INDEX	External - Market Research Agency	% Bank A mentiones out of all brand mentions, survey based***
			Brand Warmth	BRAND_WARMTH_INDEX	External - Market Research Agency	Average response across three survey questions****

**The data provider Market Research Agency collect data 24 hours a day for 7 days a week and based on an extensive database of cost associated to certain marketing activity they collect, categorise and estimate money spend on each media activity by product, media, brand, etc. For example, for press activity all key newspapers are reviewed and based on the day of the advertisement (which influences the amount of copies sold), the page and size a cost is estimated for that day for the advertiser institution split by product. All these spends then collected and aggregated up by product/campaign type and media type.

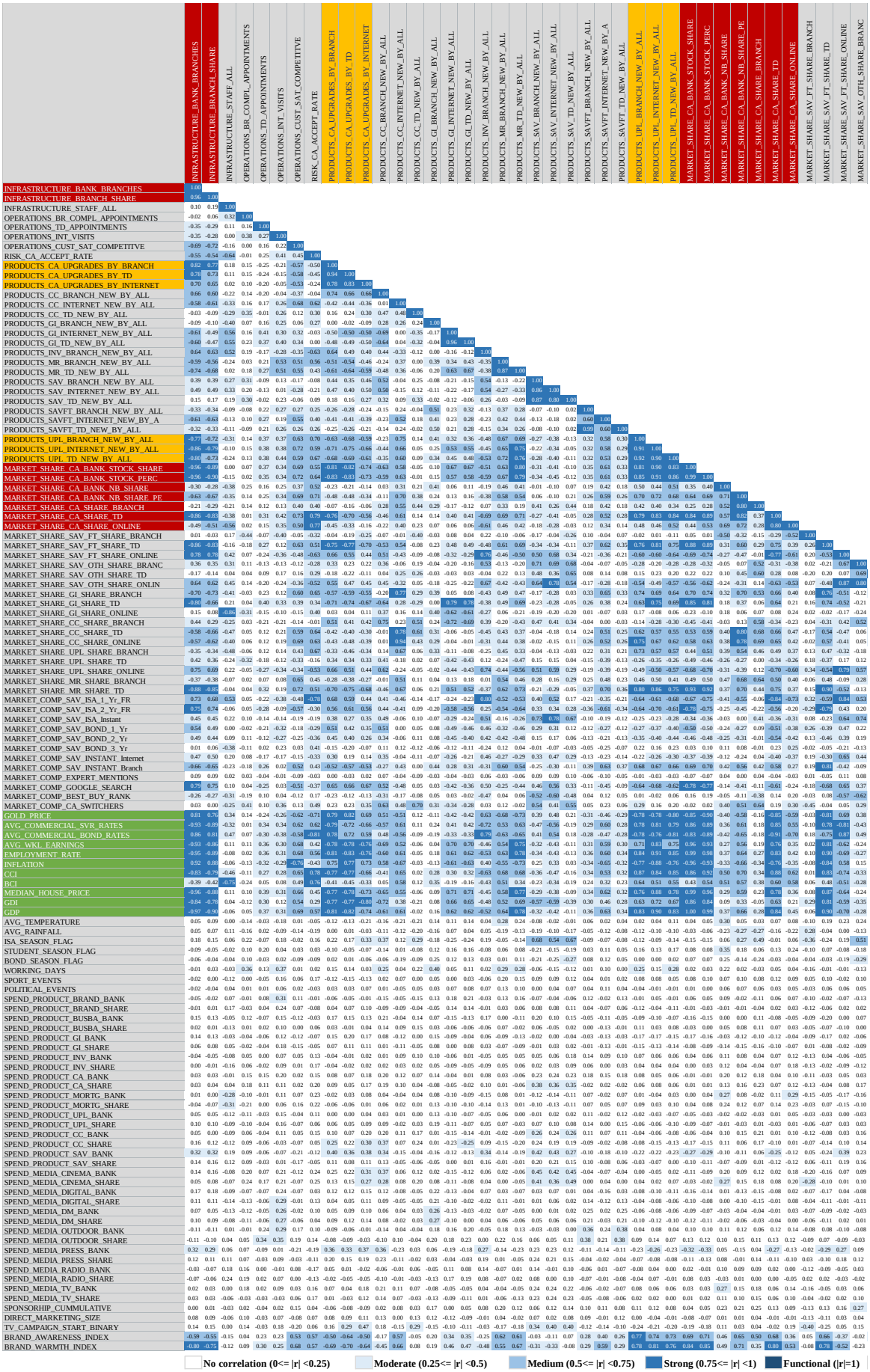
***Weekly flat phasing of monthly data

****Combination of three survey question: How thurstworthy Bank A is on a scale from 1 to 5 (1 = not, 5 = fully)? How fair Bank A is on a scale from 1 to 5? How likely you would recomment Bank A is on a scale from 1 to 5?

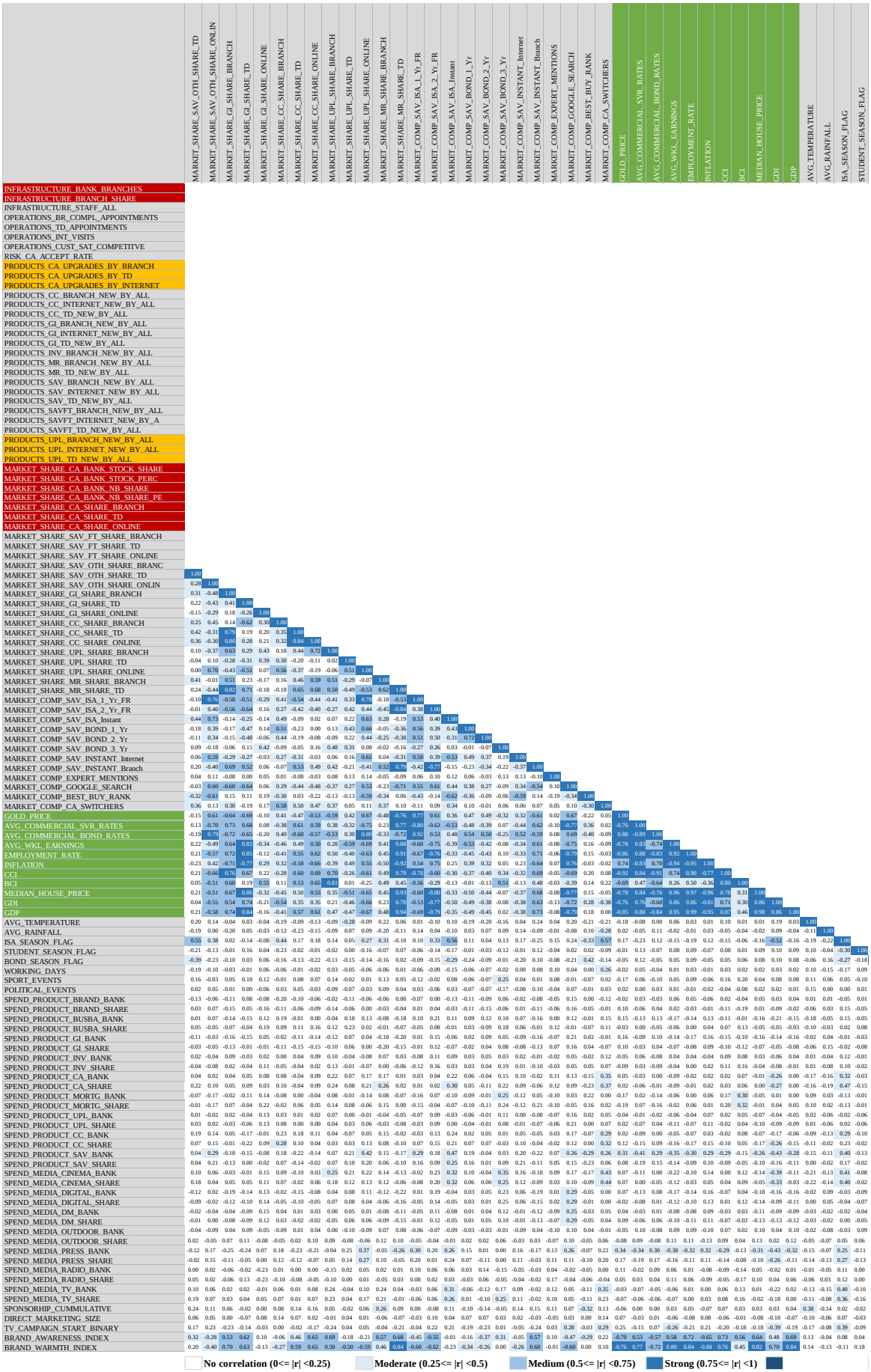
Level 1	Level 2	Level 3	Level 4	Level 5 (Variable name)	Source	Methodology
Group 4: External Factors	Economic Environment	GDP	GDP	GDP	https://www.ons.gov.uk/	Gross Domestic Product-chained volume: Seasonally adjusted £m
		Employment	Employment rate	EMPLOYMENT_RATE	https://www.ons.gov.uk/	Total weekly employed / total working age people
			Employed All	ALL_EMPLOYED	https://www.ons.gov.uk/	Phased weekly stock of monthly total number of employed people
			Employed Female	FEMALE_EMPLOYED	https://www.ons.gov.uk/	Phased weekly stock of monthly total number of employed women
			Employed Male	MALE_EMPLOYED	https://www.ons.gov.uk/	Phased weekly stock of monthly total number of employed men
			Unemployment rate	UNEMPLOYMENT_RATE	https://www.ons.gov.uk/	Phased weekly stock of monthly total number of unemployed people
		Income	Average Weekly Earning	AVG_WKL_EARNINGS	https://www.ons.gov.uk/	Average weekly total earnings
			Gross disposable income	DISPONSABLE_INCOME	https://www.ons.gov.uk/	Average weekly income remaining after deduction of taxes
		Interest rates	SVR rates	AVG_COMMERCIAL_SVR_RATES	http://www.bankofengland.co.uk/	Standard variable rate mortgage to households not seasonally adjusted
			Bond rates	AVG_COMMERCIAL_BOND_RATES	http://www.bankofengland.co.uk/	Fixed rate bond deposits from households not seasonally adjusted
		Prices	Inflation	INFLATION	https://www.ons.gov.uk/	Consumer Prices Index (% change)
			Median house price	MEDIAN_HOUSE_PRICE	https://www.ons.gov.uk/	Weekly phasing of monthly median house prices in England & Wales
			Gold Price	GOLD_PRICE	http://www.bankofengland.co.uk/	Weekly Phasing of monthly average gold price per ounce £
		Confidence Index	Concusmer Confidence Index	CCI	https://data.oecd.org/	Amplitude adjusted, Long-term average = 100, Mar 2012 – Oct 2016
			Business Confidence Index	BCI	https://data.oecd.org/	Amplitude adjusted, Long-term average = 100, Mar 2012 – Oct 2016
	Weather	Temperature and Rainfall	Average Temperature	AVG_TEMPERATURE	http://www.metoffice.gov.uk/	Weekly average England & Wales temperarture (celsius)
			Average Rainfall (mm)	AVG_RAINFALL	http://www.metoffice.gov.uk/	Weekly total England & Wales precipitation (mm)
	Seasonality	Product Seasons	ISA Season	ISA_SEASON_FLAG	Internal - Management Information System	Weekly binary flag of active season (at least 1 day = 1, none = 0)
			Student season	STUDENT_SEASON_FLAG	Internal - Management Information System	Weekly binary flag of active season (at least 1 day = 1, none = 0)
			Bond Season	BOND_SEASON_FLAG	Internal - Management Information System	Weekly binary flag of active season (at least 1 day = 1, none = 0)
	Events	Sport and Political events	Working Days	WORKING_DAYS	Internal - Management Information System	Weekly number of working days (Weekday=1, Saturday=0.5, Sunday=0)
			Sport Events	SPORT_EVENTS	External - Market Research Agency	Weekly total number of distinctive national/international sport events
			Political Events	POLITICAL_EVENTS	External - Market Research Agency	Weekly total number of distinctive national/international political events

10.14751/SZIE.2017.056

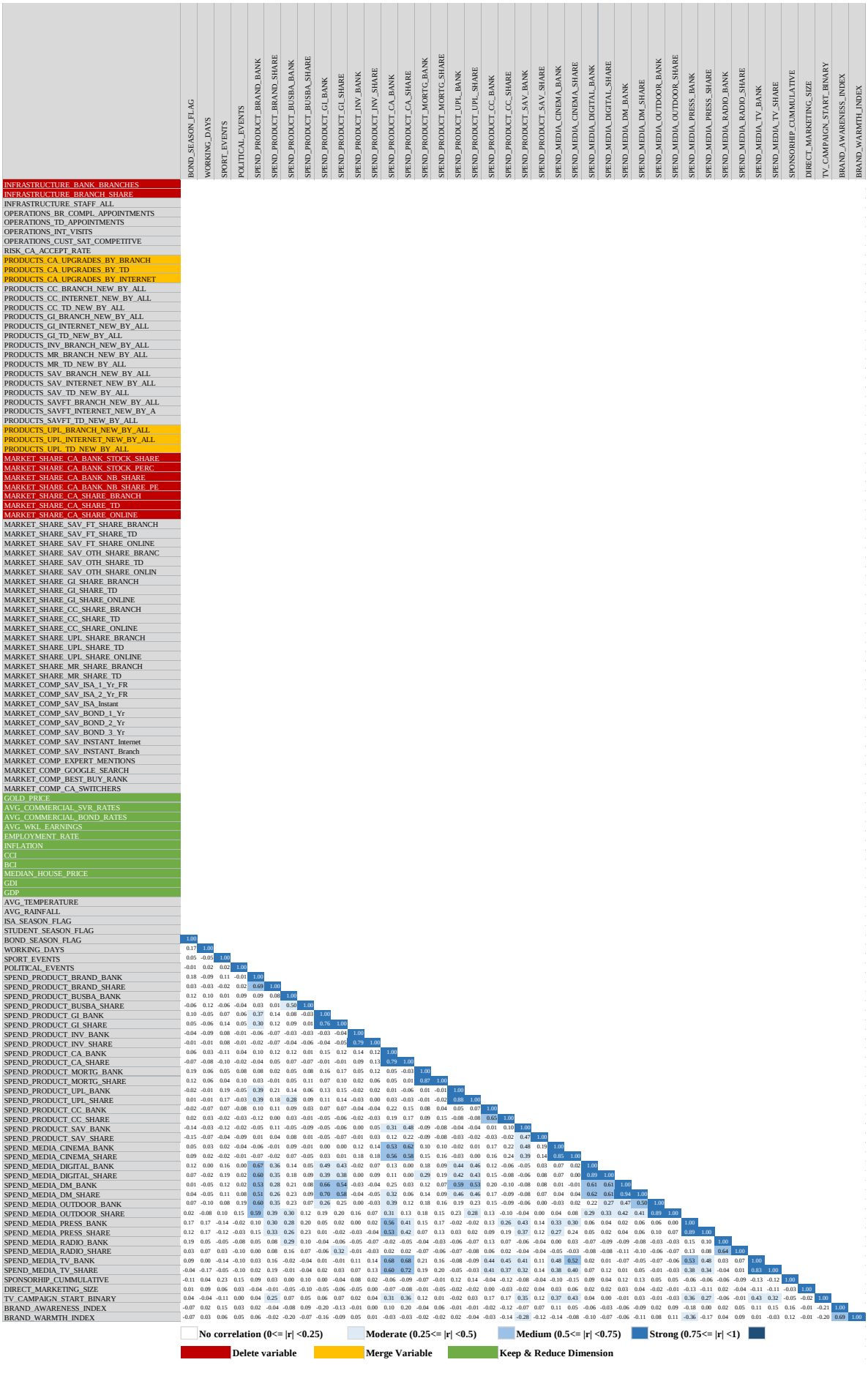
A6. Detailed Correlation Matrix (nVAR=121 nOBS=205)



Correlation Matrix Continued (2/3)



Correlation Matrix Continued (3/3)



A7. Correlation Matrix Scatter Plot (Macroeconomic variables)

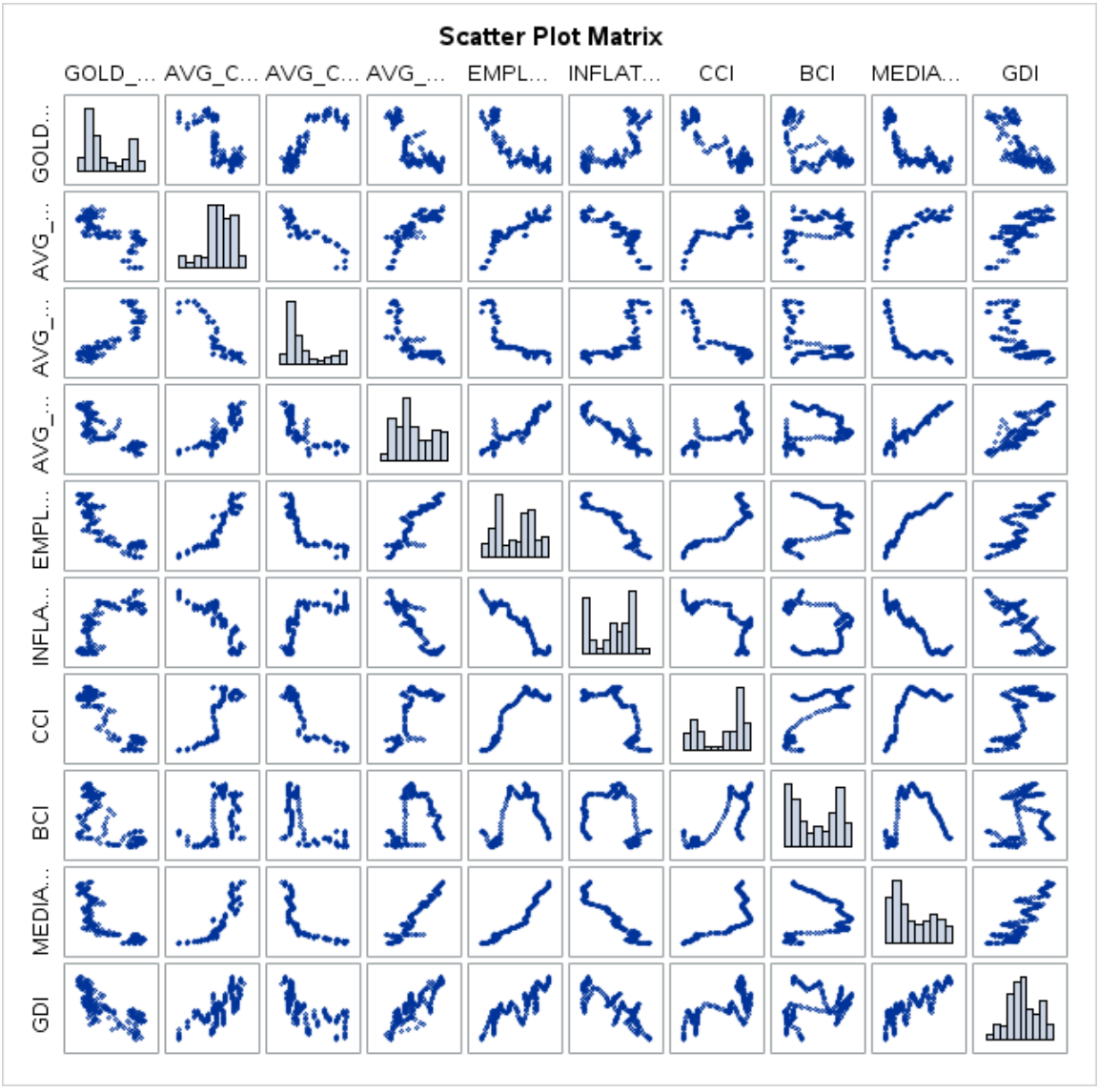
```
/* Correlation Matrix - Scatter Plots*/
ODS graphics on;

proc corr data=SASUSER.PHD_SIMULATION_DATAMART_3a
          plots(maxpoints=none)=matrix(hist nvar=all);
run;
ODS graphics off;
```

The CORR Procedure

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
GOLD_PRICE	205	859.70303	127.62931	176239	699.37600	1106
AVG_COMMERCIAL_SVR_RATES	205	4.40278	0.11283	902.57000	4.10000	4.58000
AVG_COMMERCIAL_BOND_RATES	205	1.63493	0.49694	335.16000	1.07000	2.76000
AVG_WKL_EARNINGS	205	479.23122	9.28631	98242	463.00000	497.00000
EMPLOYMENT_RATE	205	72.36424	1.09248	14835	70.50000	74.12000
INFLATION	205	1.64098	1.12332	336.40000	-0.10000	3.50000
BCI	205	100.90234	0.72657	20685	99.92950	102.01710
CCI	205	100.68479	1.65365	20640	97.86913	102.55840
MEDIAN_HOUSE_PRICE	205	207735	15411	42585638	188620	238652
GDI	205	296579	13035	60798618	267336	322806
GDP	205	442190	12396	90648896	424367	462968



A8. Factor Analysis SAS codes and outputs (Macroeconomic variables)

Bartlett's Test for sphericity

```
/* Bartlett's Test for sphericity */

PROC FACTOR DATA=SASUSER.PHD_SIMULATION_DATAMART_3a
            HEYWOOD
            METHOD=ML;
            VAR GOLD_PRICE
                AVG_COMMERCIAL_SVR_RATES
                AVG_COMMERCIAL_BOND_RATES
                AVG_WKL_EARNINGS
                EMPLOYMENT_RATE
                INFLATION
                CCI
                MEDIAN_HOUSE_PRICE
                GDI
                GDP;
RUN;
```

Convergence criterion satisfied.

Significance Tests Based on 205 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	45	5069.3092	<.0001
HA: At least one common factor			
H0: 5 Factors are sufficient	5	71.7070	<.0001
HA: More factors are needed			

Factor Analysis (Macroeconomic variables)

```
/* Factor Analysis of Macroeconomic variables */

PROC FACTOR DATA=SASUSER.PHD_SIMULATION_DATAMART_3a
            SIMPLE
            CORR
            MSA
            METHOD=PRIN
            PRIORS=SMC
            SCREE
            ROTATE=VARIMAX
            FLAG=.70
            NFactors=3
            SCORE;
            VAR GOLD_PRICE
                AVG_COMMERCIAL_SVR_RATES
                AVG_COMMERCIAL_BOND_RATES
                AVG_WKL_EARNINGS
                EMPLOYMENT_RATE
                INFLATION
                CCI
                MEDIAN_HOUSE_PRICE
                GDI
                GDP;
RUN;
```

Kaiser's Measure of Sampling Adequacy (Macroeconomic variables)

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.86392821		
GOLD_PRICE	AVG_COMMERCIAL_SVR_RATES	AVG_COMMERCIAL_BOND_RATES
0.90673502	0.83437380	0.81244889

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.86392821			
AVG_WKL_EARNINGS	EMPLOYMENT_RATE	INFLATION	CCI
0.92427407	0.91960958	0.87480890	0.79516499

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.86392821		
MEDIAN_HOUSE_PRICE	GDI	GDP
0.81379163	0.91144309	0.86543488

Prior Communality Estimates (Macroeconomic variables)

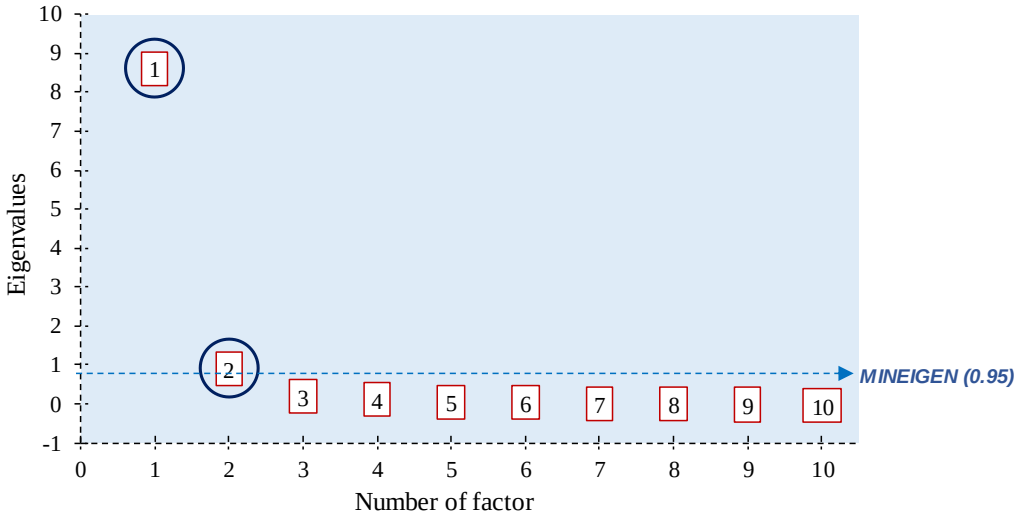
Prior Communality Estimates: SMC Total = 9.53141684 Average = 0.95314168			
GOLD_PRICE	AVG_COMMERCIAL_SVR_RATES	AVG_COMMERCIAL_BOND_RATES	AVG_WKL_EARNINGS
0.92466478	0.92909319	0.95716793	0.95838298

Prior Communality Estimates: SMC Total = 9.53141684 Average = 0.95314168					
EMPLOYMENT_RATE	INFLATION	CCI	MEDIAN_HOUSE_PRICE	GDI	GDP
0.98978247	0.9689237	0.9815119	0.99623432	0.8284972	0.9971584

Initial factor extraction (Macroeconomic variables)

Eigenvalues of the Reduced Correlation Matrix: Total = 9.83141682 Average = 0.98314168				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.604012	7.99237063	0.8752	0.8752
2	0.91164137	0.42378035	0.0927	0.9679
3	0.18786102	0.07385684	0.0191	0.9870
4	0.11400418	0.06437784	0.0116	0.9986
5	0.04962634	0.02507044	0.0050	1.0036
6	0.0245559	0.02589405	0.0025	1.0061
7	-0.00133815	0.00399269	-0.0001	1.0060
8	-0.00533085	0.01271968	-0.0005	1.0055
9	-0.01805053	0.01751391	-0.0018	1.0036
10	-0.03556444		-0.0036	1.0000

2 factors will be retained by the MINEIGEN criterion.



Factor Pattern				
	Factor1		Factor2	
GOLD_PRICE	-89	*	28	
AVG_COMMERCIAL_SVR_RATES	91	*	-10	
AVG_COMMERCIAL_BOND_RATES	-88	*	39	
AVG_WKL_EARNINGS	94	*	25	
EMPLOYMENT_RATE	99	*	6	
INFLATION	-93	*	-27	
CCI	91	*	-36	
MEDIAN_HOUSE_PRICE	96	*	25	
GDI	86	*	16	
GDP	99	*	9	

Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.7 are flagged by an '*'.

Final Communality Estimates: Total = 9.403514			
AVG_WKL_EARNINGS	EMPLOYMENT_RATE	INFLATION	CCI
0.94294971	0.98142427	0.94755576	0.96020821

Final Communality Estimates: Total = 9.403514		
MEDIAN_HOUSE_PRICE	GDI	GDP
0.98839232	0.78497972	0.99045305

Factor Rotation (Macroeconomic variables)

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix		
	1	2
1	0.75057	0.66079
2	0.66079	-0.75057

Rotated Factor Pattern				
	Factor1		Factor2	
GOLD_PRICE	-38		-80	*
AVG_COMMERCIAL_SVR_RATES	61		68	
AVG_COMMERCIAL_BOND_RATES	-30		-88	*
AVG_WKL_EARNINGS	87	*	43	
EMPLOYMENT_RATE	78	*	44	
INFLATION	-88	*	-41	
CCI	34		87	*
MEDIAN_HOUSE_PRICE	89	*	35	
GDI	75	*	35	
GDP	80	*	49	

Printer values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.7 are flagged by an '*'.

Variance Explained by Each Factor	
Factor1	Factor2
5.1141925	4.1014609

Scoring Coefficients Estimated by Regression

Squared Multiple Correlations of the Variables with Each Factor	
Factor1	Factor2
0.98746528	0.97945039

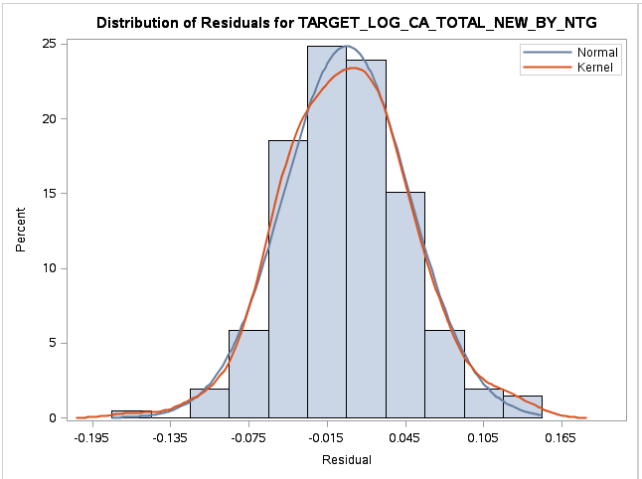
Standardized Scoring Coefficients		
	Factor1	Factor2
GOLD_PRICE	-0.083325	-0.008762
AVG_COMMERCIAL_SVR_RATES	0.17419575	-0.0509776
AVG_COMMERCIAL_BOND_RATES	0.4784672	-0.6164728
AVG_WKL_EARNINGS	0.09967946	0.09730327
EMPLOYMENT_RATE	0.11825719	0.1647683
INFLATION	-0.1978955	0.2587817
CCI	-0.3531435	0.55400101
MEDIAN_HOUSE_PRICE	1.25372406	-1.3608607
GDI	-0.0053994	0.04514136
GDP	-0.339314	0.91416695

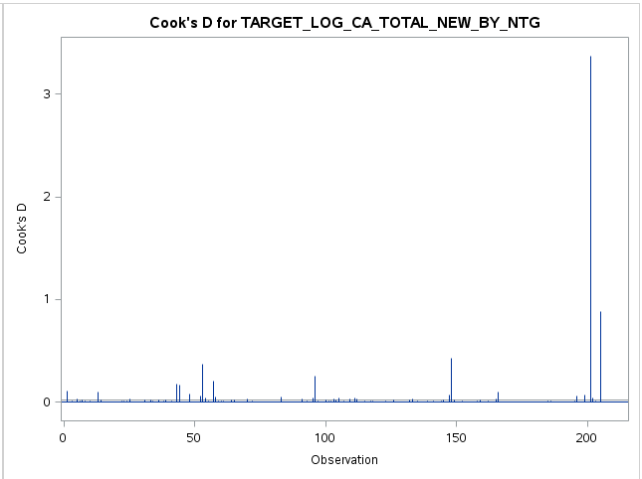
A9. Initial 'Full' Model Diagnostics

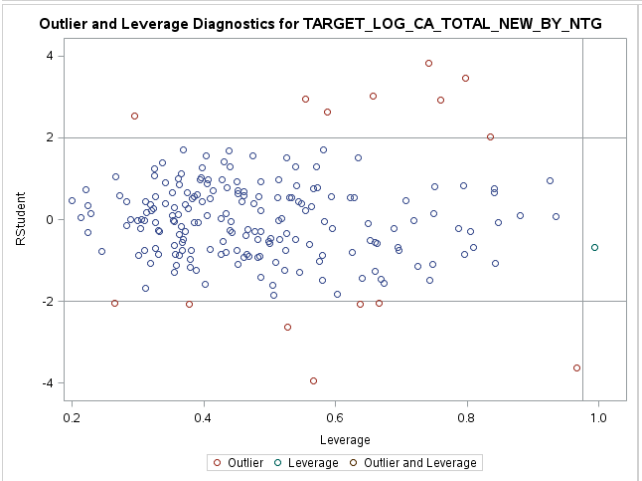
```
PROC REG DATA=SASUSER.PHD_SIMULATION_DATAMART_FULL_MODEL
        PLOTS(ONLY)=ALL;
     Linear_Regression_Model: MODEL TARGET_CA_TOTAL_NEW_BY_NTG = _ALL_
        /       SELECTION=ALL
        STB SS1 SS2 CORRB COVB CLB
        ALPHA=0.05
        COLLIN COLLINOINT TOL VIF SPEC ACOV DW;
RUN;
```

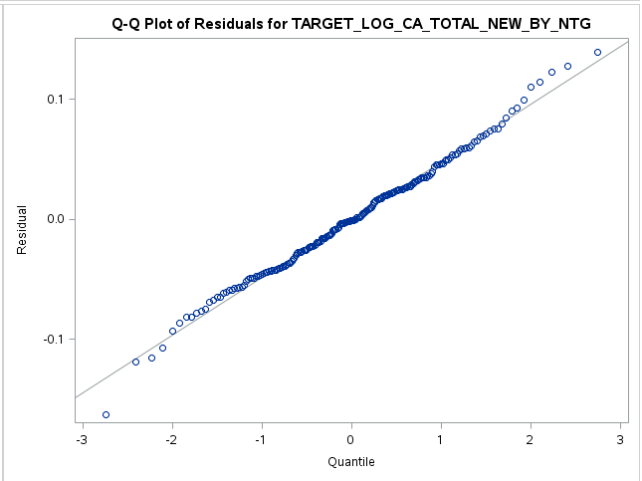
Linear Regression Results

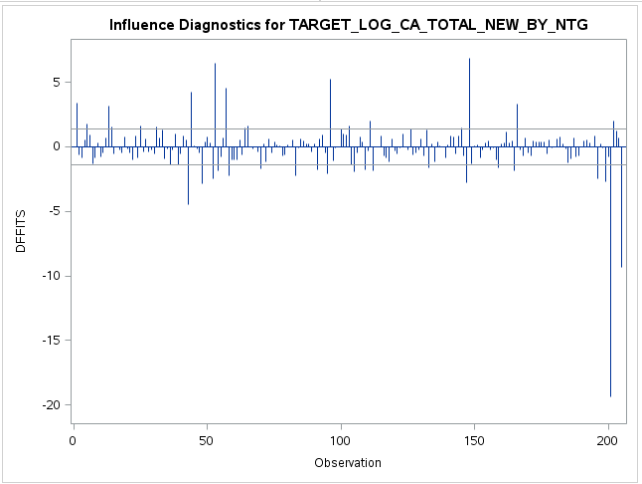
The REG Procedure
Model: Linear_Regression_Model

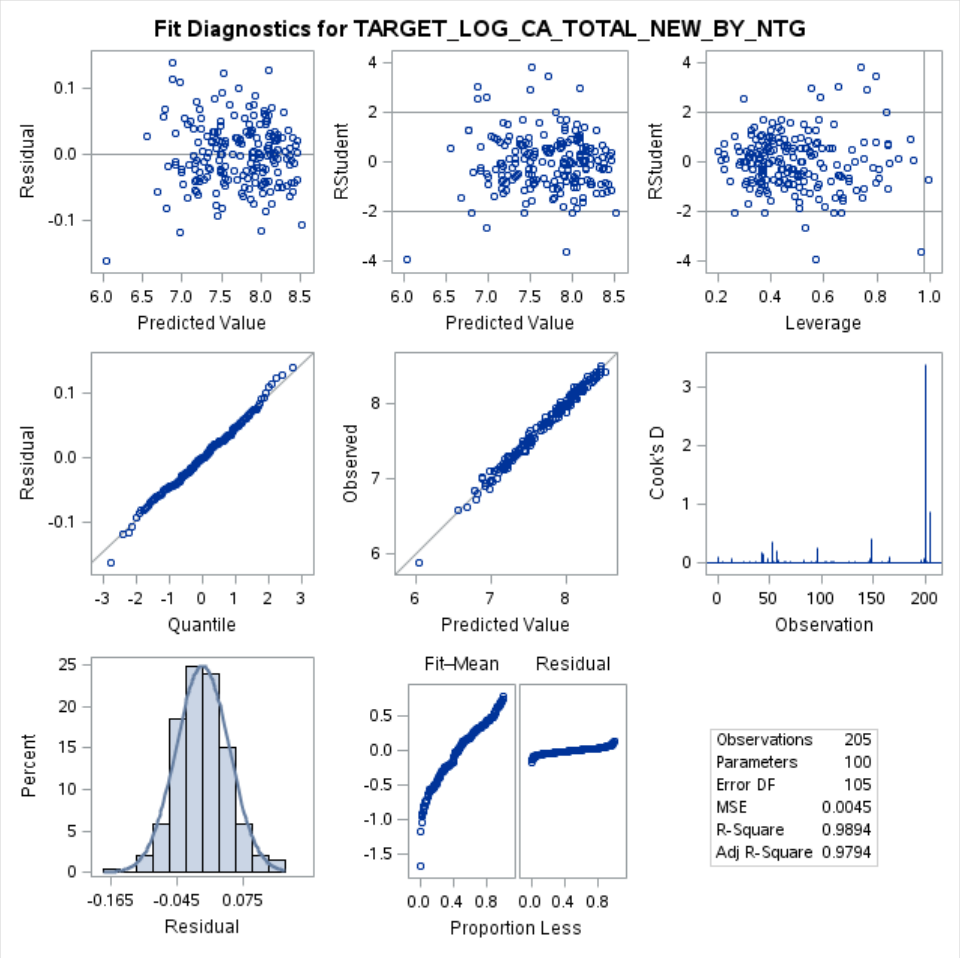


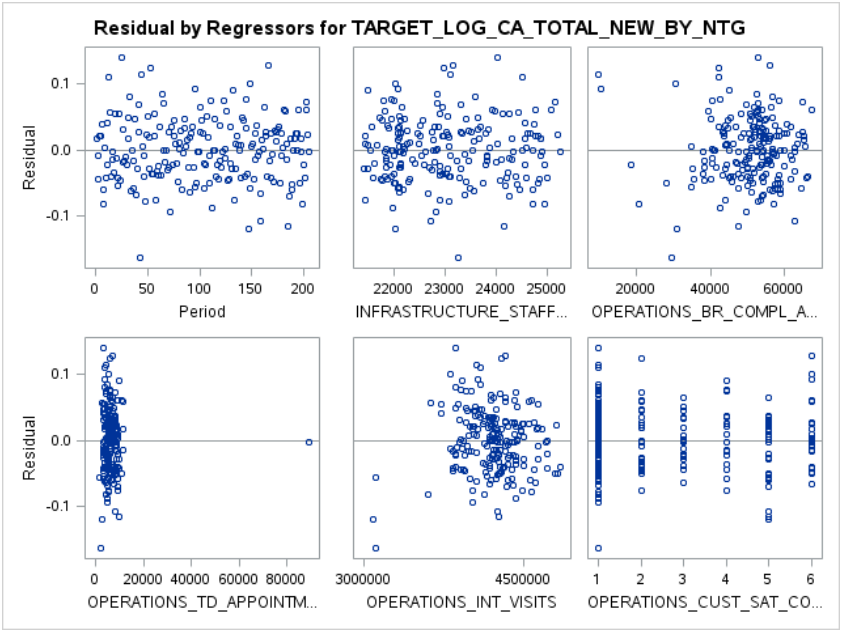


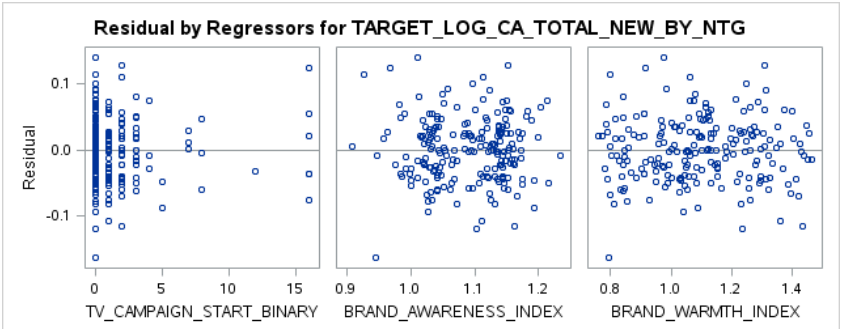


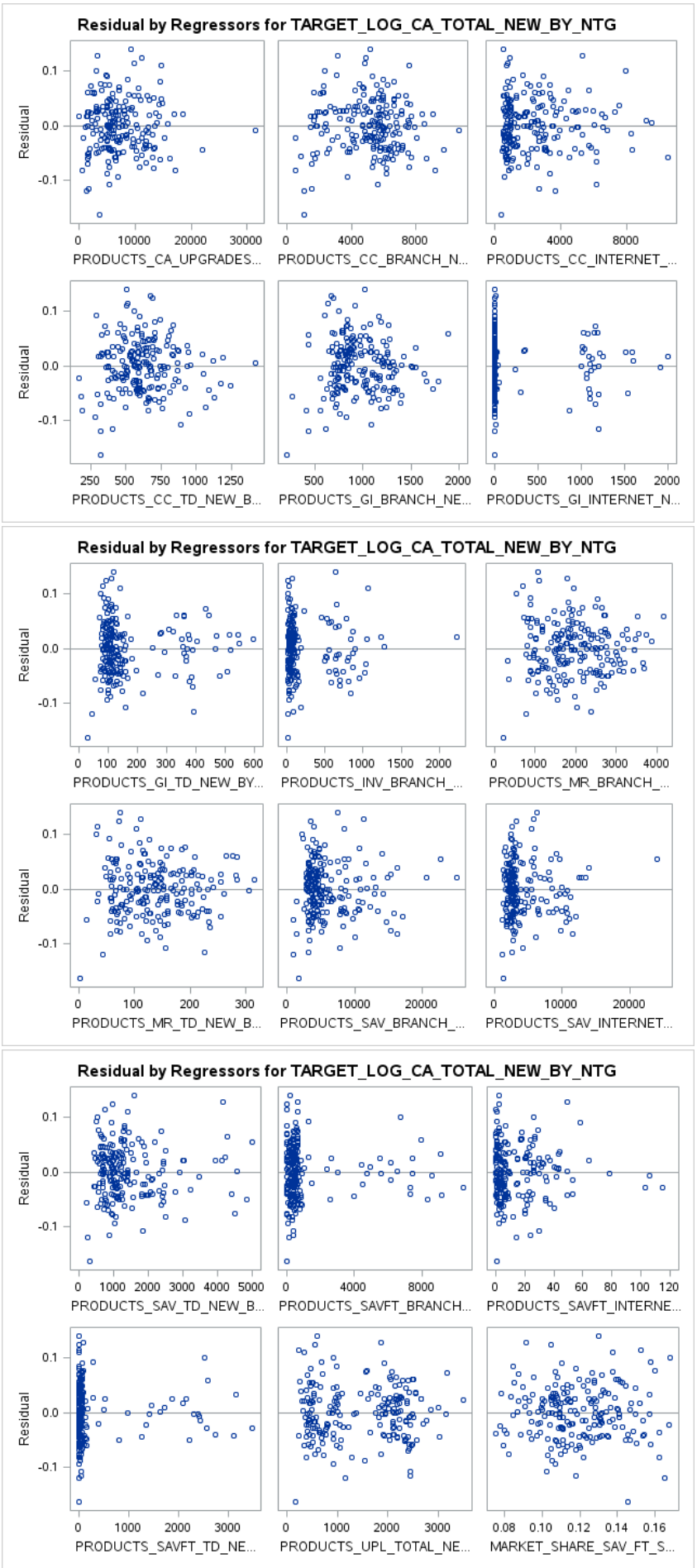


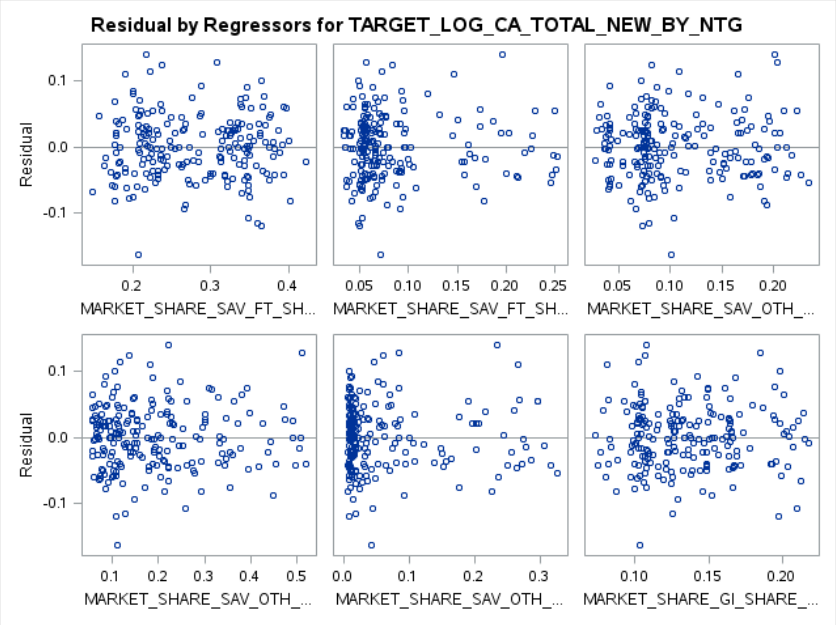


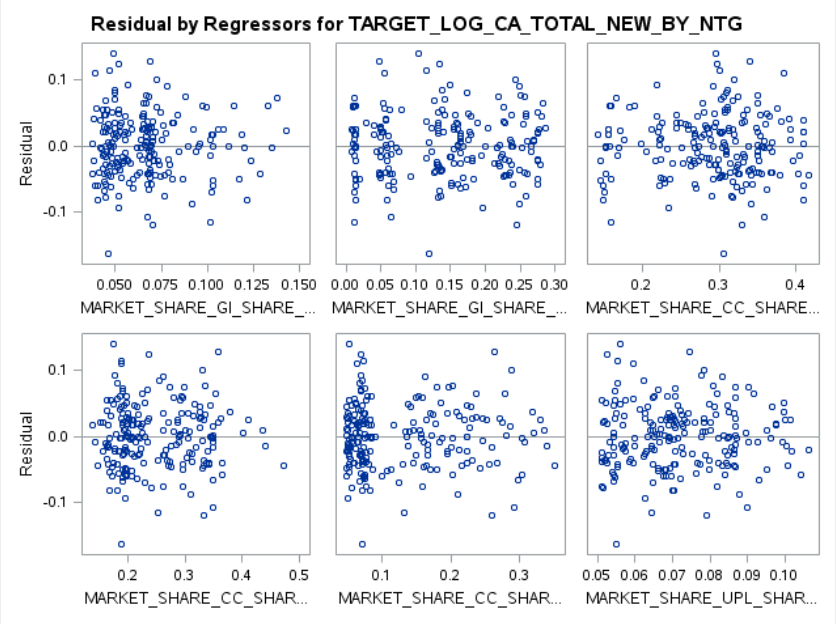

Fit Diagnostics for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Observed
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 205
Parameters 100
Error DF 105
MSE 0.0045
R-Square 0.9894
Adj R-Square 0.9794

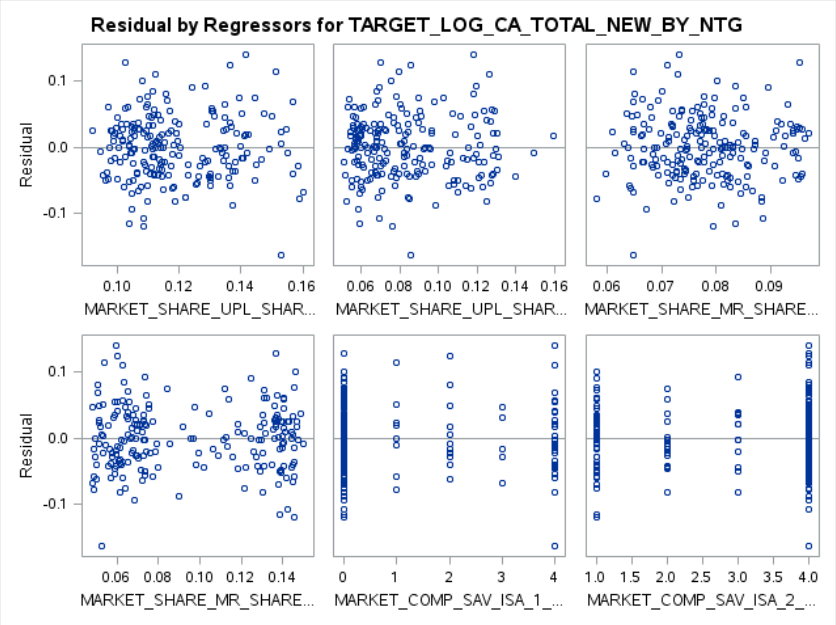

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
Period
INFRASTRUCTURE_STAFF...
OPERATIONS_BR_COMPL_A...
OPERATIONS_TD_APPOINTM...
OPERATIONS_INT_VISITS
OPERATIONS_CUST_SAT_CO...

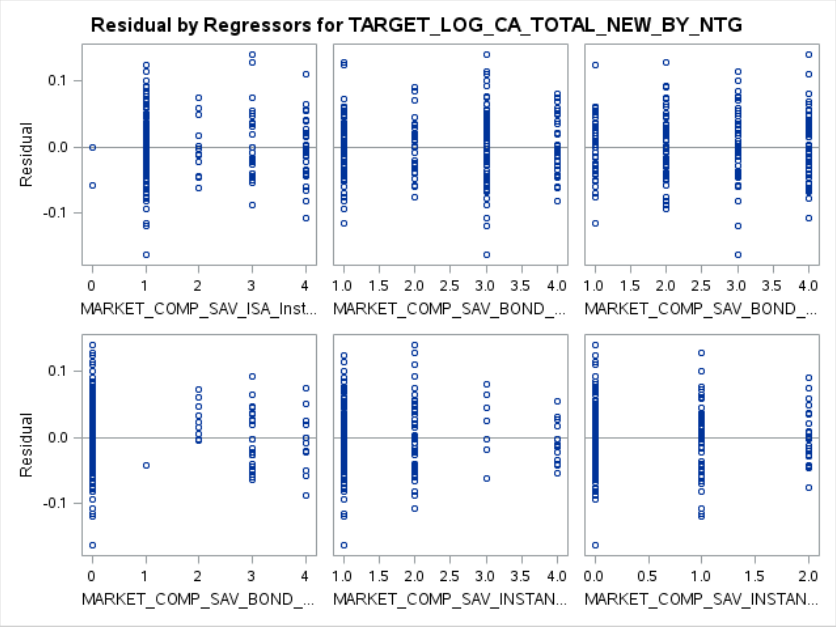

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
TV_CAMPAIGN_START_BINARY
BRAND_AWARENESS_INDEX
BRAND_WARMTH_INDEX

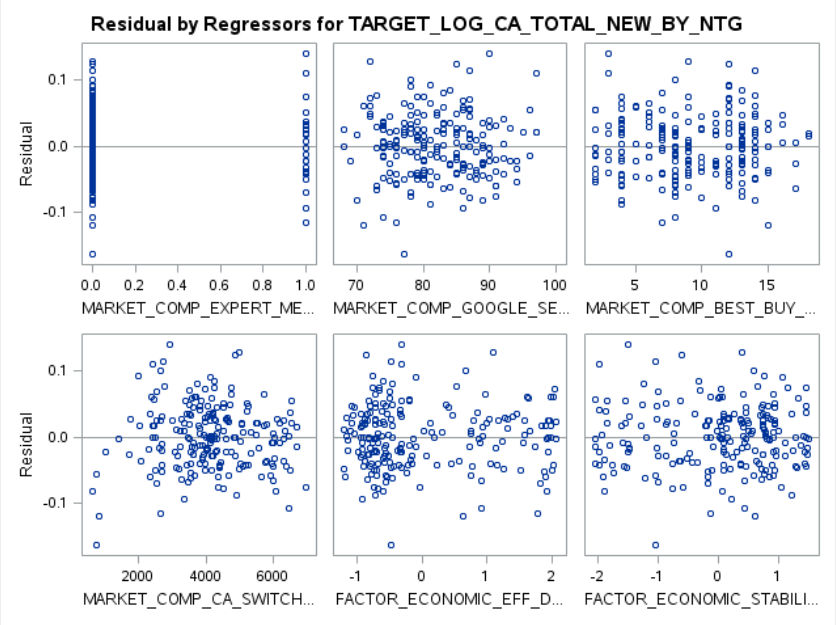

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
PRODUCTS_CA_UPGRADES...
PRODUCTS_CC_BRANCH_N...
PRODUCTS_CC_INTERNET_...
PRODUCTS_CC_TD_NEW_B...
PRODUCTS_GI_BRANCH_NE...
PRODUCTS_GI_INTERNET_N...
Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
PRODUCTS_GI_TD_NEW_BY...
PRODUCTS_INV_BRANCH_...
PRODUCTS_MR_BRANCH_...
PRODUCTS_MR_TD_NEW_B...
PRODUCTS_SAV_BRANCH_...
PRODUCTS_SAV_INTERNET...
Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
PRODUCTS_SAV_TD_NEW_B...
PRODUCTS_SAVFT_BRANCH...
PRODUCTS_SAVFT_INTERNE...
PRODUCTS_SAVFT_TD_NE...
PRODUCTS_UPL_TOTAL_NE...
MARKET_SHARE_SAV_FT_S...

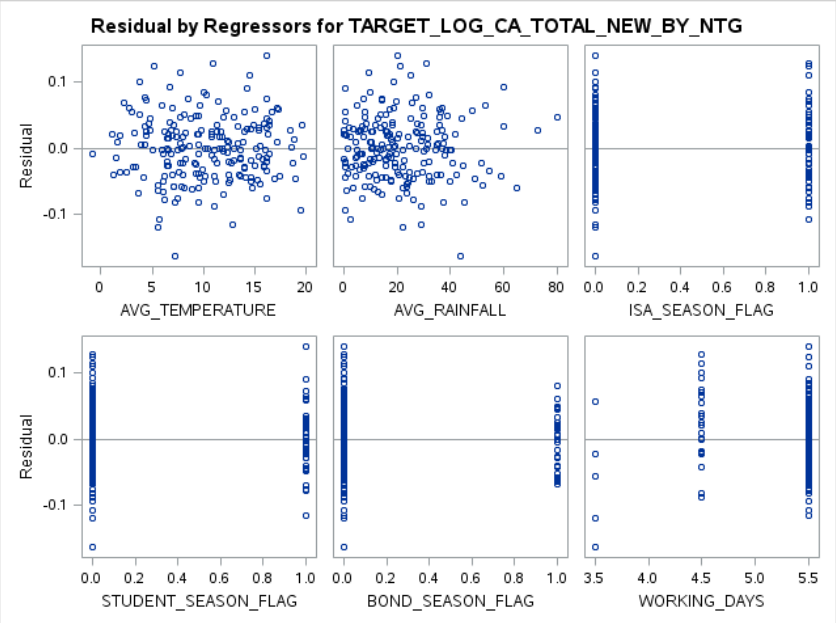

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
MARKET_SHARE_SAV_FT_SH...
MARKET_SHARE_SAV_FT_SH...
MARKET_SHARE_SAV_OTH_...
MARKET_SHARE_SAV_OTH_...
MARKET_SHARE_SAV_OTH_...
MARKET_SHARE_GI_SHARE_...

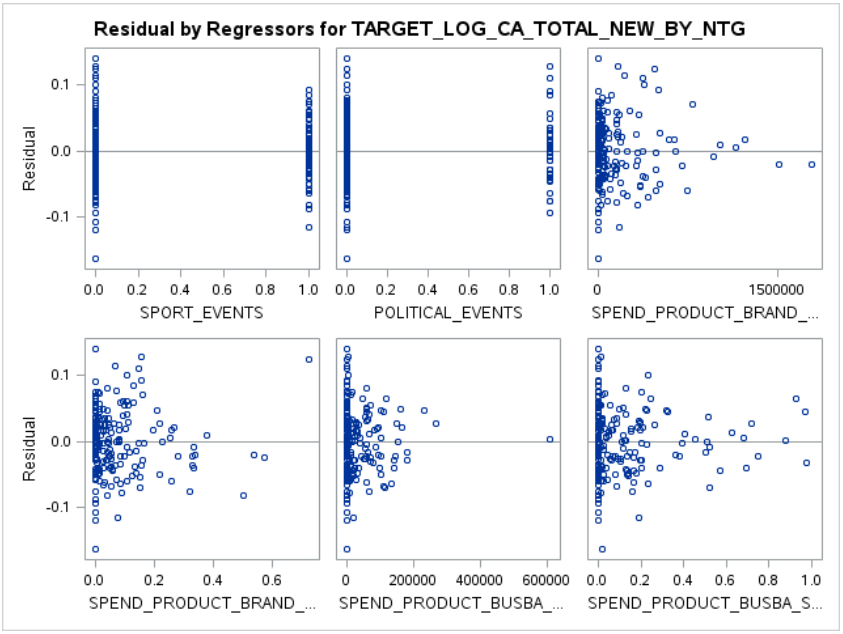

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
MARKET_SHARE_GI_SHARE_...
MARKET_SHARE_GI_SHARE_...
MARKET_SHARE_CC_SHARE...
MARKET_SHARE_CC_SHAR...
MARKET_SHARE_CC_SHAR...
MARKET_SHARE_UPL_SHAR...

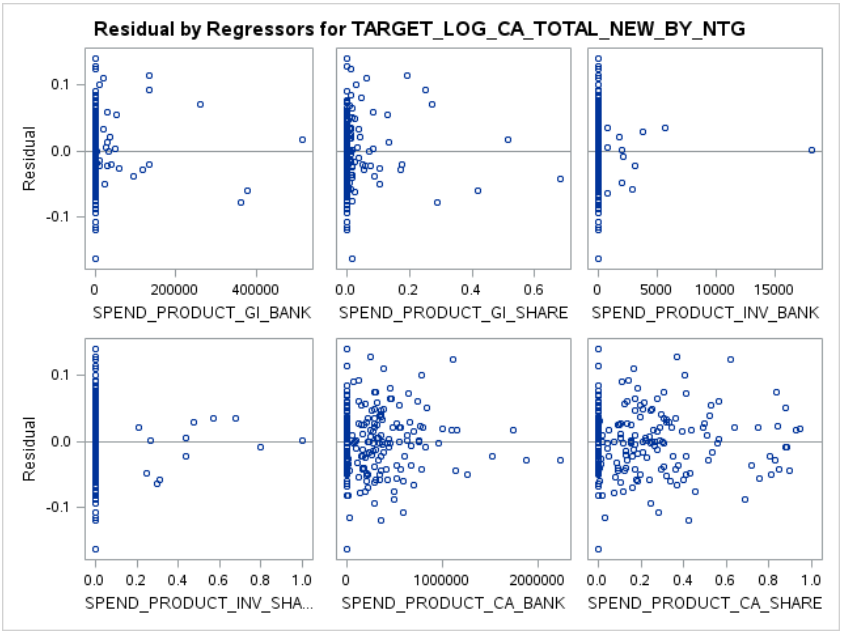

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
MARKET_SHARE_UPL_SHAR...
MARKET_SHARE_UPL_SHAR...
MARKET_SHARE_MR_SHARE...
MARKET_SHARE_MR_SHARE...
MARKET_COMP_SAV_ISA_1_...
MARKET_COMP_SAV_ISA_2_...

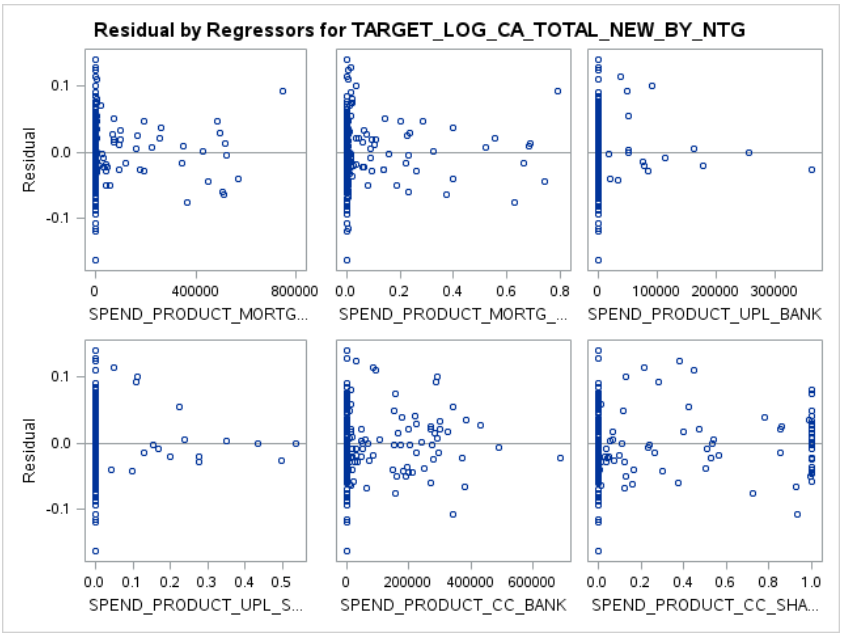

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
MARKET_COMP_SAV_ISA_Inst...
MARKET_COMP_SAV_BOND_...
MARKET_COMP_SAV_BOND_...
MARKET_COMP_SAV_BOND_...
MARKET_COMP_SAV_INSTAN...
MARKET_COMP_SAV_INSTAN...

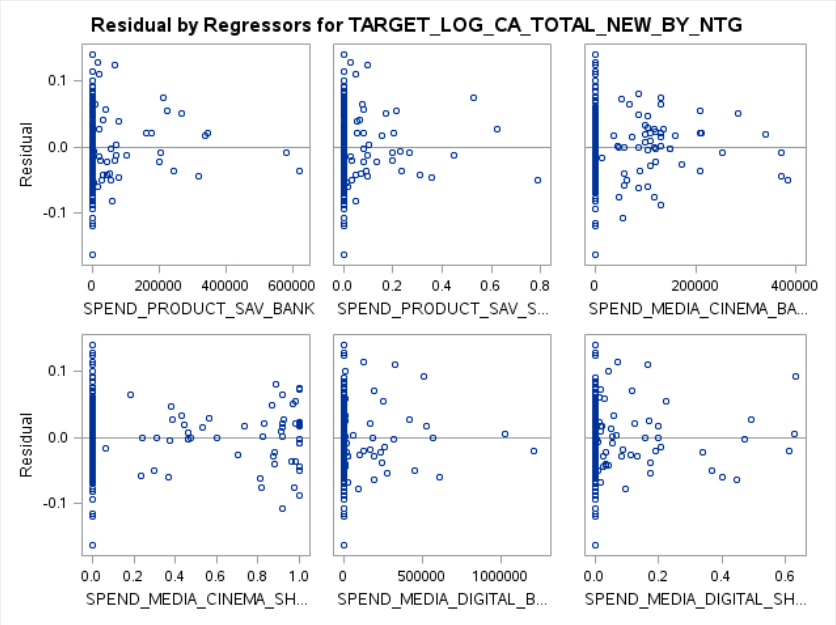

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
MARKET_COMP_EXPERT_ME...
MARKET_COMP_GOOGLE_SE...
MARKET_COMP_BEST_BUY_...
MARKET_COMP_CA_SWITCH...
FACTOR_ECONOMIC_EFF_D...
FACTOR_ECONOMIC_STABILI...

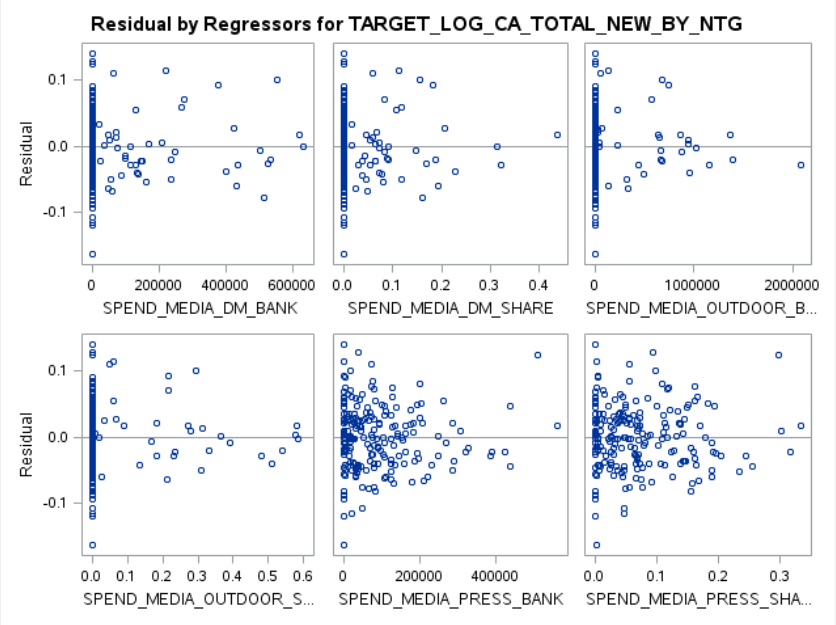

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
AVG_TEMPERATURE
AVG_RAINFALL
ISA_SEASON_FLAG
STUDENT_SEASON_FLAG
BOND_SEASON_FLAG
WORKING_DAYS

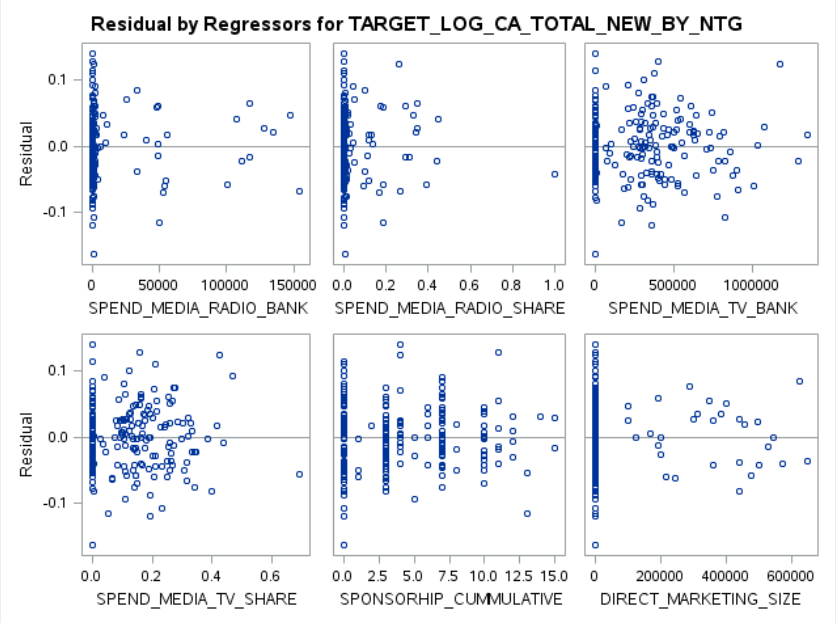

Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPORT_EVENTS
POLITICAL_EVENTS
SPEND_PRODUCT_BRAND_...
SPEND_PRODUCT_BRAND_...
SPEND_PRODUCT_BUSBA_...
SPEND_PRODUCT_BUSBA_S...


Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPEND_PRODUCT_GI_BANK
SPEND_PRODUCT_GI_SHARE
SPEND_PRODUCT_INV_BANK
SPEND_PRODUCT_INV_SHA...
SPEND_PRODUCT_CA_BANK
SPEND_PRODUCT_CA_SHARE


Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPEND_PRODUCT_MORTG...
SPEND_PRODUCT_MORTG_...
SPEND_PRODUCT_UPL_BANK
SPEND_PRODUCT_UPL_S...
SPEND_PRODUCT_CC_BANK
SPEND_PRODUCT_CC_SHA...


Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPEND_PRODUCT_SAV_BANK
SPEND_PRODUCT_SAV_S...
SPEND_MEDIA_CINEMA_BA...
SPEND_MEDIA_CINEMA_SH...
SPEND_MEDIA_DIGITAL_B...
SPEND_MEDIA_DIGITAL_SH...


Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPEND_MEDIA_DM_BANK
SPEND_MEDIA_DM_SHARE
SPEND_MEDIA_OUTDOOR_B...
SPEND_MEDIA_OUTDOOR_S...
SPEND_MEDIA_PRESS_BANK
SPEND_MEDIA_PRESS_SHA...


Residual by Regressors for TARGET_LOG_CA_TOTAL_NEW_BY_NTG
Residual
SPEND_MEDIA_RADIO_BANK
SPEND_MEDIA_RADIO_SHARE
SPEND_MEDIA_TV_BANK
SPEND_MEDIA_TV_SHARE
SPONSORHIP_CUMMULATIVE
DIRECT_MARKETING_SIZE

```
/** Normality Test - FULL Model**/
goptions reset=all;
proc univariate data=PHD_CA_CUSTOMER_FULL_MODEL normal;
 var residual;
 qqplot residual / normal(mu=est sigma=est);
run;
```

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.991876	Pr < W	0.3112
Kolmogorov-Smirnov	D	0.051221	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.055079	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.330211	Pr > A-Sq	>0.2500

```
/** Calculate Median **/
proc means data=WORK.Normality_Step1
     noprint;
 var predicted;
 output out=WORK.HETEROSCED_TEST_STEP2
        median=median;
 run;

 /** Capture Median as Parameter **/
proc sql noprint;
     select median into: median from WORK.HETEROSCED_TEST_STEP2;
 quit;
 %put &median.;

 /** Split Residuals to two equal groups based on the median of the predicted value
     (Group 1 is where the Predicted value is greater than or equal to the median
      Group 2 is where the Predicted value is less than the median) **/

PROC SQL;
    CREATE TABLE WORK.HETEROSCED_TEST_STEP2 AS SELECT
            (CASE WHEN t1.predicted >= &median. THEN 'GROUP1' ELSE 'GROUP2' END) AS GROUPS,
            t1.residual
       FROM WORK.NORMALITY_STEP1 t1
       ORDER BY t1.predicted;
 QUIT;

proc ttest data=WORK.HETEROSCED_TEST_STEP2;
     title "Two sample t-test example";
     class GROUPS;
     var residual;
 run;

PROC SQL;
 DROP TABLE WORK.HETEROSCED_TEST_STEP1;
 DROP TABLE WORK.HETEROSCED_TEST_STEP2;
 QUIT;
```

Two sample t-test example

The TTEST Procedure

Variable: residual (Residual)

GROUPS	N	Mean	Std Dev	Std Err	Minimum	Maximum
GROUP1	102	3.7847	97.2376	9.6280	-305.1	205.5
GROUP2	103	-3.7480	101.6	10.0072	-237.0	383.1
Diff (1-2)		7.5327	99.4342	13.8897		

GROUPS	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
GROUP1		3.7847	-15.3146	22.8840	97.2376	85.4787	112.8
GROUP2		-3.7480	-23.5973	16.1013	101.6	89.3334	117.7
Diff (1-2)	Pooled	7.5327	-19.8539	34.9194	99.4342	90.6310	110.1
Diff (1-2)	Satterthwaite	7.5327	-19.8483	34.9137			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	203	0.54	0.5882
Satterthwaite	Unequal	202.77	0.54	0.5881

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	102	101	1.09	0.6622

A10. 'Model 1' Model Diagnostics

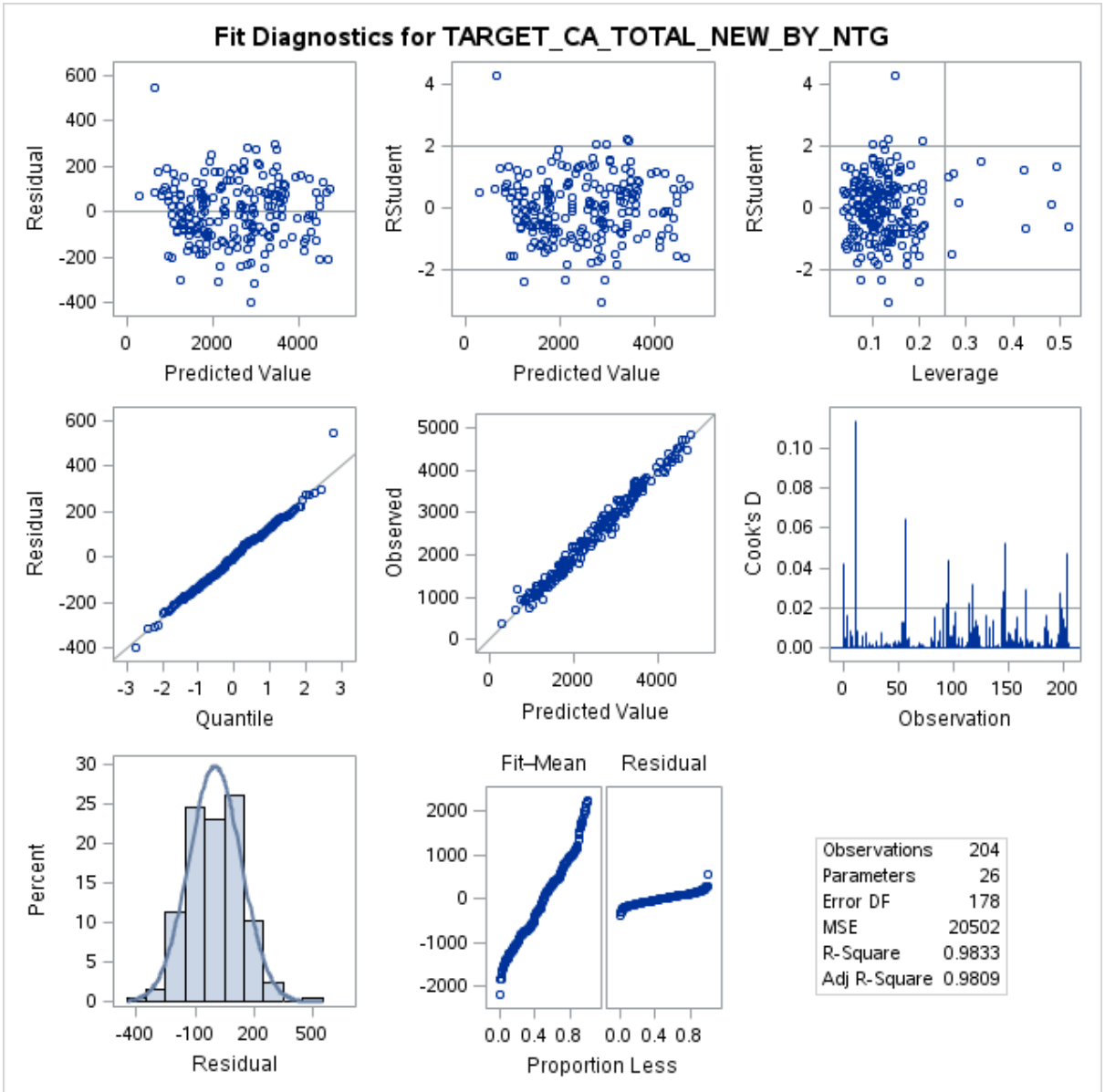
Model 1 – Initial Stepwise iteration

Stepwise Selection: Step 43

Variable PRODUCTS_CC_TD_NEW_BY_ALL Entered: R-Square = 0.9833 and C(p) = 39.4625

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	214450310	8578012	418.40	<.0001
Error	178	3649341	20502		
Corrected Total	203	218099651			



Durbin-Watson D	1.497
Number of Observations	204
1st Order Autocorrelation	0.248

Generated by the SAS System ('SASMain', Linux) on January 14, 2017 at 10:06:44 PM

10.14751/SZIE.2017.056

Model 1 – Stepwise iteration with Lag 1

Linear Regression Results

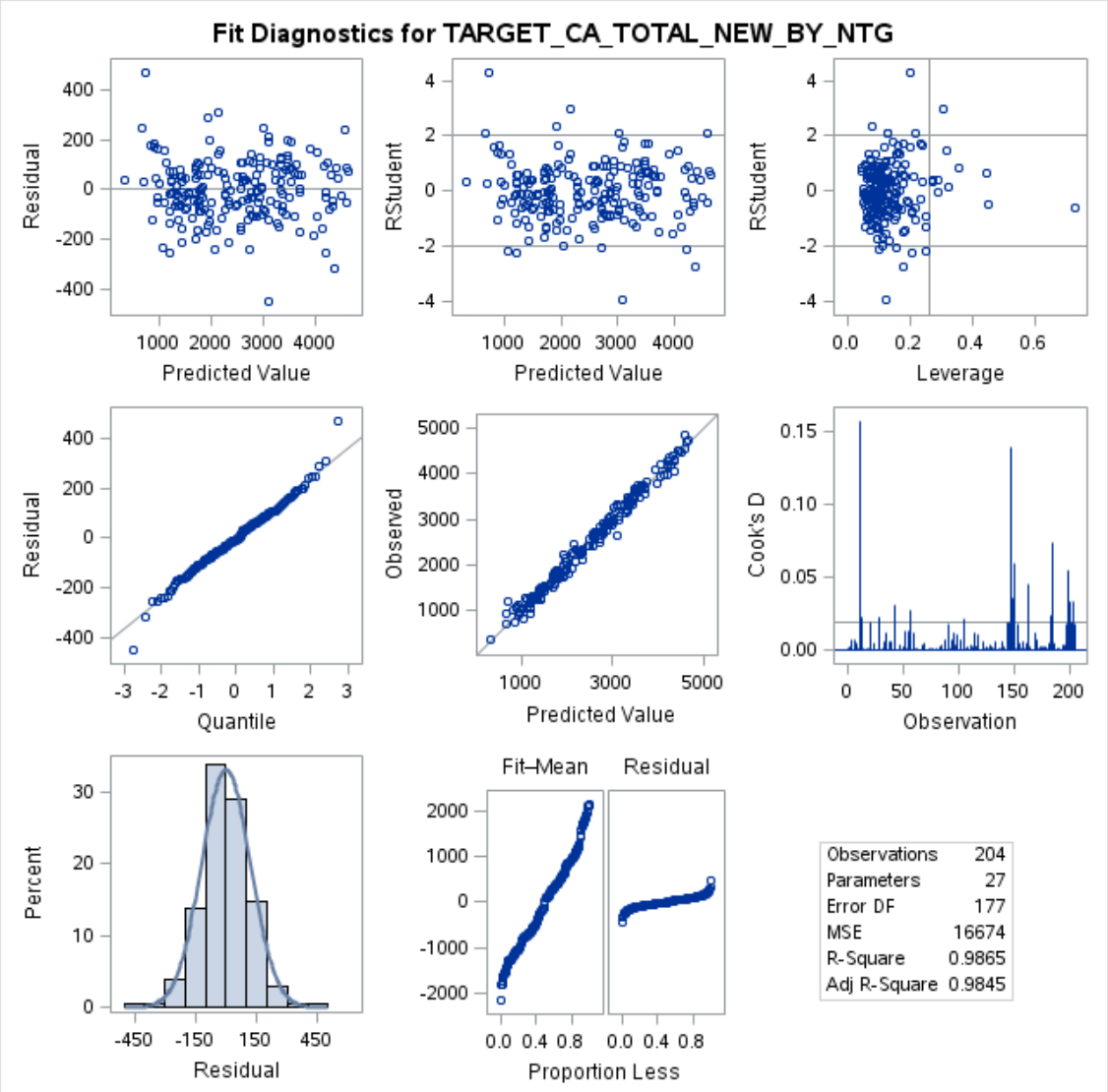
The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: TARGET_CA_TOTAL_NEW_BY_NTG

Number of Observations Read	204
Number of Observations Used	204

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	215148325	8274936	496.27	<.0001
Error	177	2951326	16674		
Corrected Total	203	218099651			

Root MSE	129.12845	R-Square	0.9865
Dependent Mean	2490.45588	Adj R-Sq	0.9845
Coeff Var	5.18493		



Durbin-Watson D	1.699
Number of Observations	204
1st Order Autocorrelation	0.145

Generated by the SAS System ('SASMain', Linux) on January 15, 2017 at 4:29:13 PM

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1387.72662	197.23460	825440	49.50	<.0001
TARGET_CA_TOTAL_NEW_BY_NTG1	0.24520	0.03393	870658	52.22	<.0001
OPERATIONS_BR_COMPL_APPOINTMENTS	-0.00944	0.00149	670169	40.19	<.0001
OPERATIONS_TD_APPOINTMENTS	0.03415	0.00992	197741	11.86	0.0007
PRODUCTS_CC_BRANCH_NEW_BY_ALL	-0.06597	0.01370	386487	23.18	<.0001
PRODUCTS_CC_INTERNET_NEW_BY_ALL	0.12828	0.01690	960554	57.61	<.0001
PRODUCTS_INV_BRANCH_NEW_BY_ALL	0.15462	0.05634	125590	7.53	0.0067
PRODUCTS_SAV_TD_NEW_BY_ALL	-0.08699	0.01799	390019	23.39	<.0001
PRODUCTS_UPL_TOTAL_NEW_BY_ALL	0.21897	0.04095	476798	28.60	<.0001
MARKET_SHARE_GI_SHARE_BRANCH	-2432.56306	675.68928	216112	12.96	0.0004
MARKET_SHARE_GI_SHARE_TD	-7642.92694	1431.09894	475580	28.52	<.0001
MARKET_SHARE_GI_SHARE_ONLINE	-530.07867	231.89307	87126	5.23	0.0234
MARKET_SHARE_CC_SHARE_ONLINE	-2059.92427	463.98676	328651	19.71	<.0001
MARKET_SHARE_UPL_SHARE_BRANCH	-8415.51507	1714.17481	401879	24.10	<.0001
MARKET_COMP_SAV_ISA_1_Yr_FR	64.46291	18.84405	195126	11.70	0.0008
MARKET_COMP_SAV_INSTANT_Internet	119.60943	16.03340	927948	55.65	<.0001
MARKET_COMP_CA_SWITCHERS	0.42349	0.02436	5037415	302.11	<.0001
FACTOR_ECONOMIC_EFF_DEMAND	317.13847	50.28049	663350	39.78	<.0001
FACTOR_ECONOMIC_STABILITY	486.66000	52.34000	1441544	86.45	<.0001
SPORT_EVENTS	71.67178	20.59062	202023	12.12	0.0006
SPEND_PRODUCT_BRAND_SHARE	-368.08547	96.22568	243983	14.63	0.0002
SPEND_PRODUCT_MORTG_BANK	0.00022678	0.00009166	102077	6.12	0.0143
SPEND_MEDIA_CINEMA_SHARE	277.29196	38.10981	882765	52.94	<.0001
SPEND_MEDIA_DIGITAL_SHARE	367.88718	101.37464	219591	13.17	0.0004
SPEND_MEDIA_RADIO_BANK	-0.00166	0.00047061	206380	12.38	0.0006
SPEND_MEDIA_RADIO_SHARE	656.98727	118.98459	508366	30.49	<.0001
BRAND_WARMTH_INDEX	238.75730	112.18856	75520	4.53	0.0347

Two sample t-test example

The TTEST Procedure

Variable: residual (Residual)

GROUPS	N	Mean	Std Dev	Std Err	Minimum	Maximum
GROUP1	102	4.0941	119.7	11.8531	-453.1	251.6
GROUP2	102	-4.0941	121.9	12.0687	-252.8	472.2
Diff (1-2)		8.1883	120.8	16.9159		

GROUPS	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
GROUP1		4.0941	-19.4192	27.6075	119.7	105.2	138.8
GROUP2		-4.0941	-28.0352	19.8469	121.9	107.1	141.4
Diff (1-2)	Pooled	8.1883	-25.1662	41.5427	120.8	110.1	133.9
Diff (1-2)	Satterthwaite	8.1883	-25.1663	41.5428			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	202	0.48	0.6289
Satterthwaite	Unequal	201.93	0.48	0.6289

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	101	101	1.04	0.8566

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.988189	Pr < W	0.0894
Kolmogorov-Smirnov	D	0.035239	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050202	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.365412	Pr > A-Sq	>0.2500

10.14751/SZIE.2017.056

A11. 'MODEL 2-4' MODEL DIAGNOSTICS

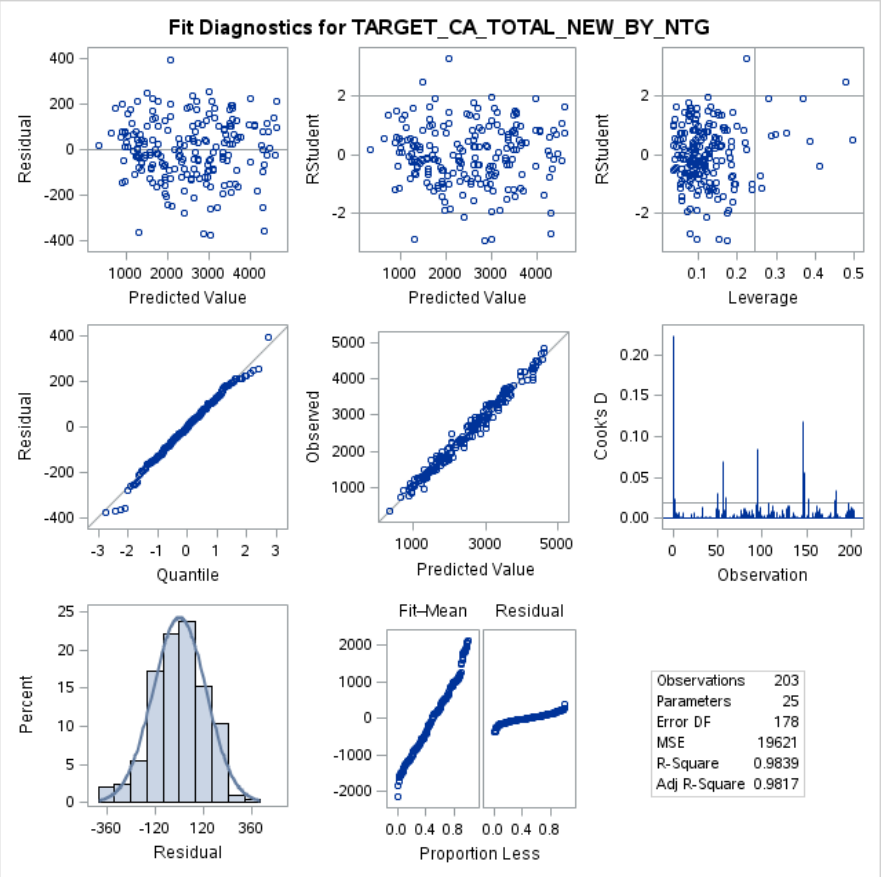
Model 2 – Stepwise iteration with Lag 1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	24	212936314	8872346	452.20	<.0001
Error	178	3492450	19621		
Corrected Total	202	216428764			

Root MSE	140.07322	R-Square	0.9839
Dependent Mean	2496.80788	Adj R-Sq	0.9817
Coeff Var	5.61009		

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Tolerance	Variance Inflation
Intercept	1	243.07459	231.38592	1.05	0.2949	0	.	0
TARGET_CA_TOTAL_NEW_BY_NTG1	1	0.18688	0.03199	5.84	<.0001	0.18686	0.08861	11.28575
OPERATIONS_BR_COMPL_APPOINTMENTS	1	-0.00998	0.00170	-5.87	<.0001	-0.08828	0.40098	2.49387
OPERATIONS_TD_APPOINTMENTS	1	0.04096	0.01125	3.64	0.0004	0.08081	0.18415	5.43049
OPERATIONS_CUST_SAT_COMPETITVE	1	37.33324	9.75983	3.83	0.0002	0.06765	0.28984	3.45018
PRODUCTS_CA_UPGRADES_BY_ALL	1	-0.05498	0.00703	-7.82	<.0001	-0.23333	0.10191	9.81210
PRODUCTS_CC_TOTAL_NEW_BY_ALL	1	0.02763	0.00807	3.42	0.0008	0.07923	0.16918	5.91103
PRODUCTS_SAV_TOTAL_NEW_BY_ALL	1	-0.01436	0.00325	-4.42	<.0001	-0.10589	0.15829	6.31748
PRODUCTS_UPL_TOTAL_NEW_BY_ALL	1	0.18467	0.04134	4.47	<.0001	0.14826	0.08232	12.14789
MARKET_SHARE_SAV_OTH_SHARE_TOTAL	1	-1064.57559	404.49769	-2.63	0.0092	-0.06136	0.16676	5.99657
MARKET_SHARE_UPL_SHARE_TOTAL	1	-12735	2025.66298	-6.29	<.0001	-0.14260	0.17621	5.67510
MARKET_SHARE_MR_SHARE_TOTAL	1	11718	2319.44271	5.05	<.0001	0.13344	0.12994	7.69593
MARKET_COMP_SAV_ISA_1_Yr_FR	1	60.09034	17.65134	3.40	0.0008	0.08308	0.15222	6.56923
MARKET_COMP_SAV_BOND_3_Yr	1	50.67525	14.19050	3.57	0.0005	0.06350	0.28669	3.48809
MARKET_COMP_SAV_INSTANT_Internet	1	115.07185	17.53402	6.56	<.0001	0.09345	0.44707	2.23679
MARKET_COMP_CA_SWITCHERS	1	0.46494	0.03037	15.31	<.0001	0.57935	0.06330	15.79840
ISA_SEASON_FLAG	1	-150.49415	43.90478	-3.43	0.0008	-0.06280	0.27010	3.70236
SPORT_EVENTS	1	66.21749	23.15781	2.86	0.0048	0.03095	0.77364	1.29259
SPEND_PRODUCT_BRAND_SHARE	1	-445.60663	105.49506	-4.22	<.0001	-0.04750	0.71675	1.39519
SPEND_PRODUCT_CA_SHARE	1	151.30542	58.15200	2.60	0.0101	0.03497	0.50173	1.99310
SPEND_PRODUCT_MORTG_BANK	1	0.00109	0.00018277	5.99	<.0001	0.13310	0.18340	5.45269
SPEND_PRODUCT_MORTG_SHARE	1	-774.24526	159.35208	-4.86	<.0001	-0.10727	0.18598	5.37702
SPEND_MEDIA_CINEMA_SHARE	1	252.92052	43.00957	5.88	<.0001	0.08688	0.41535	2.40762
SPEND_MEDIA_DIGITAL_BANK	1	0.00021797	0.00008065	2.70	0.0075	0.03113	0.68325	1.46360
SPEND_MEDIA_RADIO_SHARE	1	564.65089	101.47480	5.56	<.0001	0.05982	0.78443	1.27480

Generated by the SAS System ('SASMain', Linux) on January 16, 2017 at 7:26:45 PM

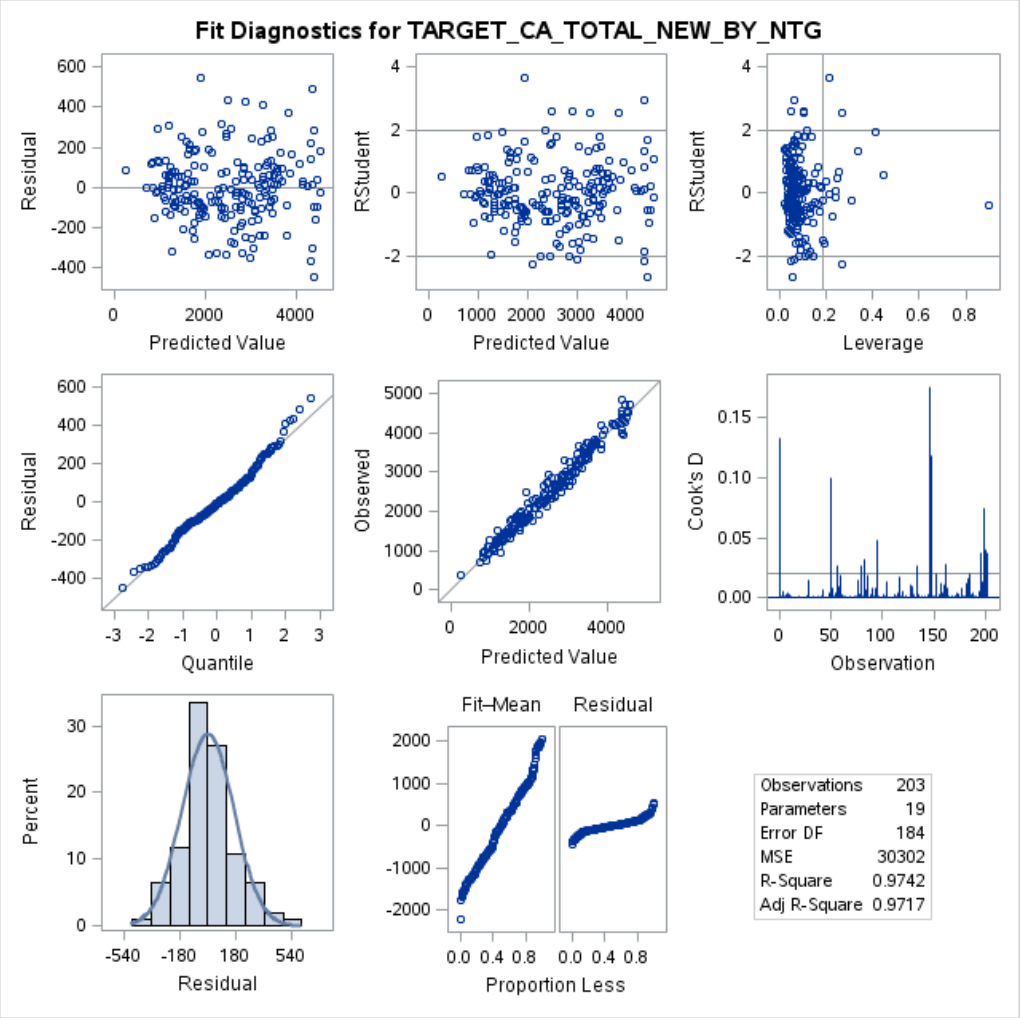


Model 3 – Stepwise iteration with Lag 1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	210853149	11714064	386.57	<.0001
Error	184	5575615	30302		
Corrected Total	202	216428764			

Root MSE	174.07543	R-Square	0.9742
Dependent Mean	2496.80788	Adj R-Sq	0.9717
Coeff Var	6.97192		

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Tolerance	Variance Inflation
Intercept	1	-153.06941	91.09315	-1.68	0.0946	0	.	0
TARGET_CA_TOTAL_NEW_BY_NTG1	1	0.28836	0.03730	7.73	<.0001	0.28834	0.10062	9.93838
OPERATIONS_BR_COMPL_APPOINTMENTS	1	-0.00618	0.00169	-3.65	0.0003	-0.05466	0.62435	1.60167
OPERATIONS_TD_APPOINTMENTS	1	0.08367	0.01130	7.41	<.0001	0.16506	0.28184	3.54814
OPERATIONS_CUST_SAT_COMPETITVE	1	51.93479	10.47947	4.96	<.0001	0.09411	0.38826	2.57556
PRODUCTS_CA_UPGRADES_BY_ALL	1	-0.04333	0.00716	-6.05	<.0001	-0.18388	0.15156	6.59821
PRODUCTS_CC_TOTAL_NEW_BY_ALL	1	0.02262	0.00917	2.47	0.0146	0.06486	0.20238	4.94111
PRODUCTS_SAV_TOTAL_NEW_BY_ALL	1	-0.00837	0.00254	-3.30	0.0012	-0.06171	0.39916	2.50523
PRODUCTS_UPL_TOTAL_NEW_BY_ALL	1	0.18148	0.04106	4.42	<.0001	0.14570	0.12887	7.75966
MARKET_COMP_SAV_BOND_1_Yr	1	-40.66149	16.41843	-2.48	0.0142	-0.04420	0.43965	2.27454
MARKET_COMP_SAV_INSTANT_Internet	1	106.11283	19.63202	5.41	<.0001	0.08618	0.55077	1.81563
MARKET_COMP_CA_SWITCHERS	1	0.34736	0.03099	11.21	<.0001	0.43283	0.09391	10.64805
BOND_SEASON_FLAG	1	85.67572	37.29920	2.30	0.0227	0.03169	0.73545	1.35971
SPEND_PRODUCT_BRAND_SHARE	1	-456.11533	120.54282	-3.78	0.0002	-0.04862	0.84784	1.17947
SPEND_PRODUCT_GI_BANK	1	-0.00101	0.00035122	-2.86	0.0047	-0.05575	0.36925	2.70816
SPEND_PRODUCT_GI_SHARE	1	1022.38889	263.35690	3.88	0.0001	0.07717	0.35432	2.82233
SPEND_PRODUCT_CA_BANK	1	0.00014007	0.00004000	3.50	0.0006	0.04661	0.79051	1.26501
SPEND_PRODUCT_MORTG_BANK	1	0.00126	0.00021740	5.79	<.0001	0.15322	0.20020	4.99489
SPEND_PRODUCT_MORTG_SHARE	1	-777.64088	190.63631	-4.08	<.0001	-0.10774	0.20069	4.98279



Model 4 – Stepwise iteration with Cochrane-Orcutt correction

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	198500774	15269290	185.55	<.0001
Error	186	15305954	82290		
Corrected Total	199	213806728			

Root MSE	286.86247	R-Square	0.9284
Dependent Mean	2491.87500	Adj R-Sq	0.9234
Coeff Var	11.51191		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Tolerance	Variance Inflation
Intercept	Intercept	1	616.99410	545.65830	1.13	0.2596	0	.	0
WORKING_DAYS		1	396.78507	47.11396	8.42	<.0001	0.16629	0.98714	1.01302
SPEND_MEDIA_CINEMA_BANK		1	0.00264	0.00037547	7.03	<.0001	0.16297	0.71701	1.39468
SPEND_MEDIA_TV_BANK		1	0.00027148	0.00008142	3.33	0.0010	0.07297	0.80363	1.24435
OPERATIONS_BR_COMPL_APPOINTMENTS		1	-0.00586	0.00294	-1.99	0.0475	-0.05111	0.58607	1.70629
OPERATIONS_INT_VISITS		1	-0.00052550	0.00010456	-5.03	<.0001	-0.13179	0.55971	1.78665
PRODUCTS_CC_BRANCH_NEW_BY_ALL		1	-0.15863	0.02269	-6.99	<.0001	-0.28258	0.23558	4.24478
PRODUCTS_CC_INTERNET_NEW_BY_ALL		1	0.23535	0.01527	15.41	<.0001	0.46634	0.42031	2.37922
MARKET_COMP_SAV_INSTANT_Branch		1	137.22561	40.70535	3.37	0.0009	0.09271	0.50888	1.96509
MARKET_COMP_GOOGLE_SEARCH		1	11.97989	5.11486	2.34	0.0202	0.07489	0.37643	2.65652
MARKET_COMP_CA_SWITCHERS		1	0.37124	0.03226	11.51	<.0001	0.45764	0.24332	4.10986
AVG_RAINFALL		1	5.96801	1.42829	4.18	<.0001	0.08839	0.86009	1.16266
MARKET_SHARE_GI_SHARE_ONLINE		1	-2635.85882	423.80060	-6.22	<.0001	-0.21747	0.31482	3.17644
FACTOR_ECONOMIC_STABILITY		1	394.41362	43.39459	9.09	<.0001	0.38086	0.21920	4.56206

Generated by the SAS System ('SASMain', Linux) on January 16, 2017 at 10:57:30 PM

Tests for Normality

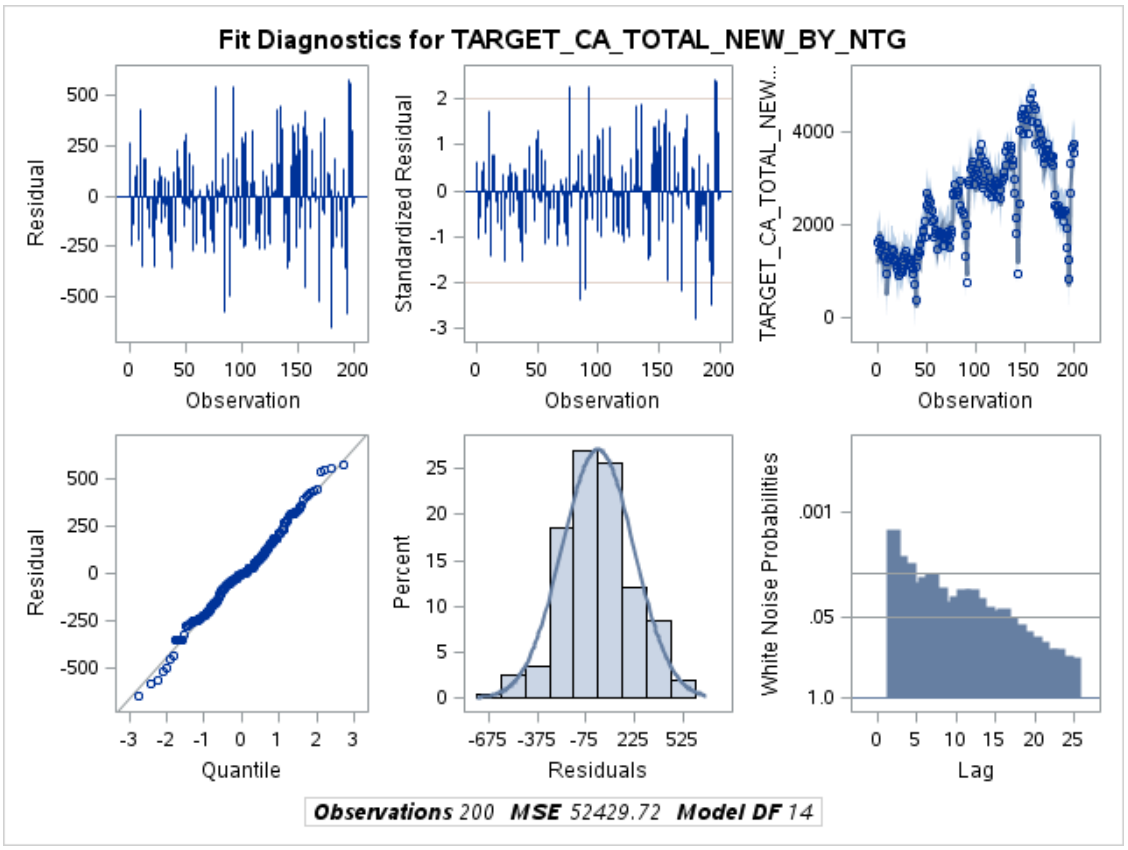
Test	Statistic		p Value	
Shapiro-Wilk	W	0.989029	Pr < W	0.1285
Kolmogorov-Smirnov	D	0.065471	Pr > D	0.0358
Cramer-von Mises	W-Sq	0.15036	Pr > W-Sq	0.0236
Anderson-Darling	A-Sq	0.809017	Pr > A-Sq	0.0376

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	99	99	1.12	0.5696

Estimates of Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	76529.8	1.000000	|********************|
1	35800.1	0.467793	|*********|



A12. Summary of the 10 Additional Final Models

Model Diagnostics of the final Champion Models

	Product Related Regression Models			
	New to Bank A Customers (via Current Account)	New to Banking Existing Customers (via CA)	Existing Customers with new Current Account	All (existing and new) customers with new CA
	Model 1: CA_NTG	Model 2: CA_NTB	Model 3: CA_EXIST	Model 4: CA_ALL
Variable	TARGET_CA_TOTAL_NEW_BY_NTG	TARGET_CA_TOTAL_NEW_BY_NTB	TARGET_CA_TOTAL_NEW_BY_EXIST	TARGET_CA_TOTAL_NEW_BY_ALL
Rsqr	0.983	0.9883	0.9381	0.9849
Rsqr adj	0.9807	0.9868	0.9304	0.9829
DW	1.544	1.575	1.201	1.367
VIF	6.29	4.13	6.17	6.24
Obs	202	202	202	202
Var	24	22	23	24
S.E. %	0.054091842	0.037460695	0.065910832	0.036508498
Shapiro-Wilk	0.0836	0.9488	0.09	0.0611
Eq. Var.	0.2401	0.0819	0.1224	0.1874

	Product Related Regression Models	
	All (existing and new) customers with new Credit C.	All (existing and new) customers with new Savings
	Model 5: CC_ALL	Model 6: SAV_ALL
Variable	TARGET_CC_TOTAL_NEW_BY_ALL	TARGET_SAV_TOTAL_NEW_BY_ALL
Rsqr	0.9036	0.9074
Rsqr adj	0.8986	0.8988
DW	0.926	1.277
VIF	1.98	4.38
Obs	202	202
Var	10	17
S.E. %	0.110988883	0.203223505
Shapiro-Wilk	0.0541	0.0432
Eq. Var.	0.0511	0.0332

	Market Share Related Regression Models			
	Current Account New Business (Campaign) Market Share	Stock Market Share Position (Current Accounts)	Credit Card New Business (Campaign) Market Share	Savings New Business (Campaign) Market Share
	Model 7: MRK_CA	Model 8: MRKS_CA	Model 9: MRK_CC	Model 10: MRK_SAV
Variable	MARKET_SHARE_NB_CA_TOTAL	MARKET_SHARE_STOCK_CA_TOTAL	MARKET_SHARE_NB_CC_TOTAL	MARKET_SHARE_NB_SAV_TOTAL
Rsqr	0.9204	0.9979	0.9493	0.9142
Rsqr adj	0.9126	0.9976	0.9444	0.9058
DW	0.897	0.933	1.128	0.781
VIF	3.89	6.48	2.74	3.57
Obs	202	202	202	202
Var	18	26	18	18
S.E. %	0.03388921	0.006509795	0.062577817	0.192217179
Shapiro-Wilk	0.0532	0.2259	0.0529	0.064
Eq. Var.	0.9514	0.4041	0.8651	0.499

Combined list of the Standardised Estimates of the additional final model

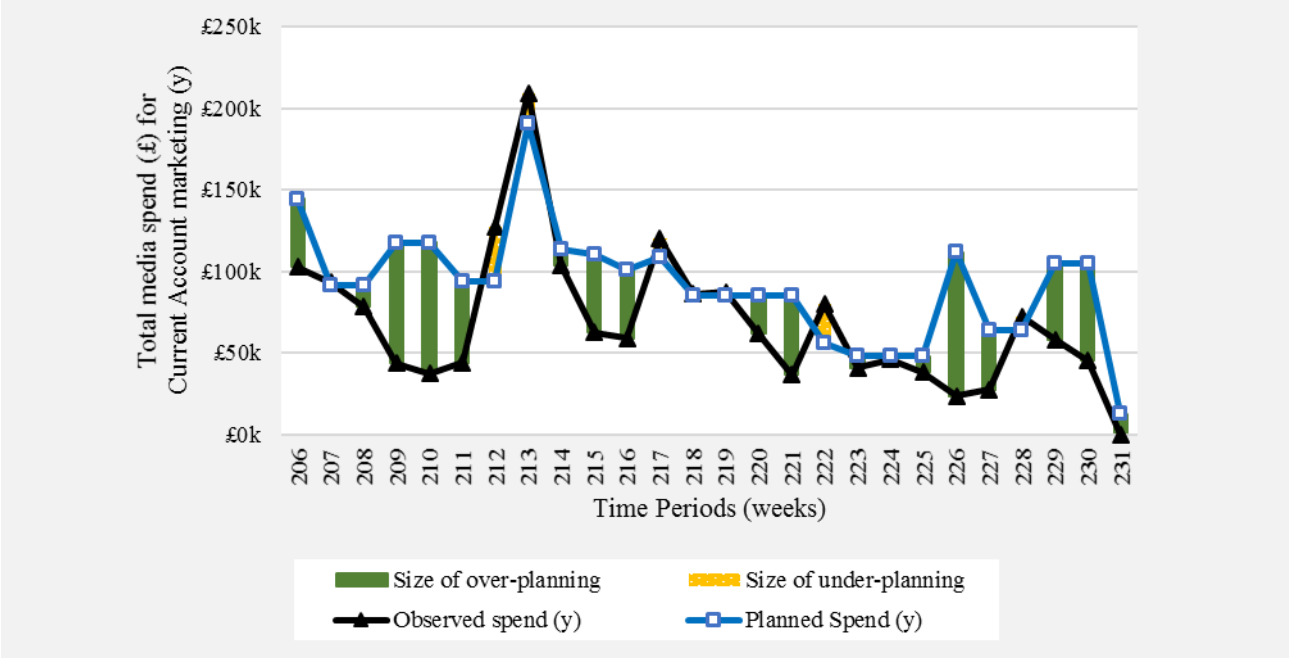
	Standardized Estimates (Beta Coefficients)									
	Product Related Regression Models						Market Share Related Regression Models			
	CA_NTG	CA_NTB	CA_EXIST	CA_ALL	CC_ALL	SAV_ALL	MRK_CA	MRK_CC	MRK_SAV	MRKS_CA
Intercept	0	0	0	0	0	0	0	0	0	0
OPERATIONS_BR_COMPL_APPOINTMENTS	-0.06221	-0.05533	0	0	0	0	0	0	0	0.03066
OPERATIONS_TD_APPOINTMENTS	0.08509	0	0	0	0	-0.12415	0.41294	-0.15529	0	-0.03634
OPERATIONS_CUST_SAT_COMPETITVE	0.10249	0	0	0	0	0	0.20005	0	0	0
PRODUCTS_CA_UPGRADES_BY_ALL	-0.17399	-0.16152	0	-0.08922	0	0	0	0	0	0.07224
PRODUCTS_SAV_TOTAL_NEW_BY_ALL	-0.13642	0	0	0	0	0	0	0	0.43162	0
PRODUCTS_UPL_TOTAL_NEW_BY_ALL	0.18471	0	0	0.10788	0	0	0	0	0	0.0995
MARKET_SHARE_UPL_SHARE_TOTAL	-0.11247	-0.06053	-0.21187	-0.10366	0.09287	0	0	0	0	0
MARKET_SHARE_MR_SHARE_TOTAL	0.11777	-0.11562	0.24444	0	0	0.29711	0.19914	0	0	0
MARKET_COMP_SAV_ISA_1_Yr_FR	0.14979	0	0.1821	0.09279	0	-0.13799	0	0.24355	0.22814	-0.03602
MARKET_COMP_SAV_INSTANT_Internet	0.1089	0	0.19806	0.11171	0	0	-0.11005	0	0.10434	0
MARKET_COMP_SAV_INSTANT_Branch	-0.09564	-0.08973	-0.14987	-0.10423	0	0	-0.15298	0	0	0
MARKET_COMP_CA_SWITCHERS	0.73853	0.82854	0.74744	0.84025	0	0.64137	0.31499	0	0	-0.069
FACTOR_ECONOMIC_EFF_DEMAND	0.15144	0	-0.4989	0	0	0	0	0	0	0.12966
FACTOR_ECONOMIC_STABILITY	0.28375	0	0.1948	0.2256	0	0	0	-0.37293	-0.26628	0.09396
ISA_SEASON_FLAG	-0.09549	0	0	0	-0.09759	0	0	0.05809	0.15391	0
WORKING_DAYS	-0.04024	0	0.07634	0	0.11668	-0.07884	-0.08672	-0.10227	0	-0.01308
SPORT_EVENTS	0.03302	0	0	0	0	0	0	-0.05497	0	0
SPEND_PRODUCT_BRAND_SHARE	-0.04906	0	-0.06116	-0.04838	0	0	0	0	0	0
SPEND_PRODUCT_MORTG_BANK	0.12139	0.05417	0	0.0462	0	0	0	0	0	0
SPEND_PRODUCT_MORTG_SHARE	-0.092	-0.03851	0	0	0	0	0	0	0	0
SPEND_PRODUCT_UPL_BANK	-0.02386	0	0	0	0	0	0	0	0	0
SPEND_MEDIA_CINEMA_SHARE	0.10709	0	0	0	0	0	0	0	0.13172	-0.04282
SPEND_MEDIA_DIGITAL_BANK	0.04122	0	0.07761	0.04826	0	0	0	0	0	0
SPEND_MEDIA_RADIO_SHARE	0.06075	0	0	0.02519	0	0.06468	0	0	0	0
INFRASTRUCTURE_STAFF_ALL		-0.29119	0	-0.20871	0	0	-0.27035	0	0	0
OPERATIONS_INT_VISITS		0.05066	0	0	0.11025	0	0	0	0	0
PRODUCTS_CC_TOTAL_NEW_BY_ALL		0.1449	0.2982	0.15585	0	0	-0.41647	0.5305	0	0
PRODUCTS_GI_TOTAL_NEW_BY_ALL		-0.09423	-0.24516	-0.16818	0	0	0	0	0	-0.02456
PRODUCTS_INV_TOTAL_NEW_BY_ALL		-0.03816	0	0	0	0.1558	0	0	-0.13801	-0.02496
MARKET_SHARE_SAV_FT_SHARE_TOTAL		-0.11304	0	0	0	0	0	0	0	0
MARKET_SHARE_CC_SHARE_TOTAL		-0.26199	-0.2822	-0.16645	0.61878	-0.36577	0.48056	0	0.31095	0
MARKET_COMP_SAV_ISA_2_Yr_FR		0.07775	-0.16727	0	-0.11004	0	0	0	0	0
STUDENT_SEASON_FLAG		-0.03352	0	0	0	0	0.21439	-0.12409	0	0
SPEND_PRODUCT_INV_BANK		0.01818	0	0	0	0	0	0	-0.05584	0
SPEND_PRODUCT_CC_BANK		0.0204	0	0	0	0	0	0.04148	0	0
SPEND_MEDIA_DIGITAL_SHARE		0.02332	0	0	0	0	0	0	0	0
SPONSORHIP_CUMMULATIVE		0.02318	0	0	0	0	0	0	0	0
BRAND_AWARENESS_INDEX		0.11107	0	0	0	0	0.08569	0	0	0
PRODUCTS_SAVFT_TOTAL_NEW_BY_ALL		0	0.10534	0.06583	0	0	0	0	0	0
MARKET_SHARE_SAV_OTH_SHARE_TOTAL		0	-0.3749	-0.18541	0	0.39125	0.25446	0	0	0
MARKET_SHARE_GI_SHARE_TOTAL		0	-0.33411	-0.22359	0	0	0	0.3847	0	0
MARKET_COMP_SAV_BOND_3_Yr		0	0.21278	0.06125	0	0	0.1536	0	0.12509	-0.05355
BOND_SEASON_FLAG		0	0.0869	0.02979	0	0	0	0	0	0
TV_CAMPAIGN_START_BINARY		0	-0.06848	0	0	0	0.09789	0	0.07787	0
BRAND_WARMTH_INDEX		0	0.1674	0.10691	0	0	0	0	0	0.02286
MARKET_COMP_SAV_BOND_2_Yr		0	0	-0.04748	0	0	0	0.10496	0	0.03909
PRODUCTS_MR_TOTAL_NEW_BY_ALL		0	0	0.05643	-0.18386	0	0	0.15969	0	0
MARKET_COMP_SAV_BOND_1_Yr		0	0	0	0.10427	0	0	0	-0.13281	-0.05878
TARGET_CA_TOTAL_NEW_BY_ALL						-0.25393	0	0	-0.15713	0.07876
MARKET_COMP_SAV_ISA_Instant						0.27659	0	0	0.10187	0
MARKET_COMP_GOOGLE_SEARCH						0.2151	0	-0.18247	0	0
SPEND_PRODUCT_SAV_BANK						-0.10242	0	0	0	0
SPEND_MEDIA_CINEMA_BANK						0.19584	0	-0.14189	-0.20296	0.02952
AVG_TEMPERATURE						-0.20292	0.07429	0	0.28106	-0.01952
SPEND_MEDIA_TV_BANK						-0.18584	0	0	0	0.02028
SPEND_MEDIA_TV_SHARE						0.08581	0	0	0	0
SPEND_PRODUCT_SAV_SHARE						0	-0.06094	0	0	0
SPEND_MEDIA_RADIO_BANK						0	0.084	0	-0.07098	-0.00797
MARKET_SHARE_CA_SHARE_TOTAL						0	0	0.41684	0	0.85201
SPEND_PRODUCT_BUSBA_SHARE						0	0	0.06104	0	0
SPEND_PRODUCT_UPL_SHARE						0	0	0.04018	0	-0.00965
SPEND_MEDIA_PRESS_SHARE						0	0	-0.06851	0	0
SPEND_PRODUCT_CA_BANK						0	0	0	-0.05609	0

A13. Comparison of the Plan, Actuals and Predictions (Simulation Figures)

Please note these are figures based on the simulation database.

Period	Planned (Cost Plan)	Observed (Actual Cost)	Forecasted (Predicted Cost)	Plan vs. Actual		Plan vs. Forecast	
				Size of over-planning	Size of under-planning	Size of over-planning	Size of under-planning
206	£144,395	£102,483	£98,694	£41,912	£0	£45,700	£0
207	£91,745	£93,103	£83,140	£0	£1,358	£8,605	£0
208	£91,745	£78,361	£76,677	£13,384	£0	£15,068	£0
209	£117,482	£43,578	£70,701	£73,903	£0	£46,780	£0
210	£117,482	£37,329	£69,845	£80,153	£0	£47,636	£0
211	£94,123	£44,089	£82,194	£50,034	£0	£11,929	£0
212	£94,123	£126,909	£113,098	£0	£32,785	£0	£18,974
213	£190,292	£209,255	£149,181	£0	£18,964	£41,111	£0
214	£113,760	£103,583	£125,620	£10,176	£0	£0	£11,860
215	£110,542	£62,589	£102,952	£47,953	£0	£7,590	£0
216	£101,182	£58,904	£73,020	£42,278	£0	£28,162	£0
217	£108,982	£120,128	£108,188	£0	£11,145	£794	£0
218	£85,382	£85,935	£48,854	£0	£553	£36,528	£0
219	£85,382	£87,227	£116,739	£0	£1,845	£0	£31,357
220	£85,382	£61,703	£64,337	£23,679	£0	£21,045	£0
221	£85,382	£36,442	£45,289	£48,940	£0	£40,093	£0
222	£55,599	£79,688	£16,555	£0	£24,088	£39,044	£0
223	£47,799	£40,638	£58,924	£7,161	£0	£0	£11,125
224	£47,799	£45,907	£46,866	£1,892	£0	£934	£0
225	£47,799	£38,024	£29,649	£9,775	£0	£18,151	£0
226	£111,798	£23,475	£49,270	£88,324	£0	£62,528	£0
227	£63,999	£27,313	£64,750	£36,686	£0	£0	£751
228	£63,999	£72,305	£55,674	£0	£8,305	£8,326	£0
229	£104,949	£58,148	£28,771	£46,801	£0	£76,178	£0
230	£104,949	£45,120	£51,922	£59,829	£0	£53,027	£0
231	£12,542	£0	£0	£12,542	£0	£12,542	£0
Total	**£2,378,615**	**£1,782,236**	**£1,830,912**	**£695,424**	**£99,045**	**£621,771**	**£74,068**
% vs. Plan		***75%***	***77%***				

Planned and Observed Current Account Media spend (Source: Compiled by author)



A14. Financial Benefit Calculation (Simulation Figures)

Applied Formula:

$$C_i = \left[PV \times \left(1 + \frac{r}{T}\right)^t\right] - PV$$

, where C_i is the compound interest (Benefit I and Benefit II columns),
PV is the present value (Difference between Plan and Forecast column, if the value <0 then it is 0)
r is the annual interest rate (ROI or profit margin, in current study calculation is done for 1-2-3-5%)
T is the number of times the interest is compounded (in current study it is weekly therefore 52)
t is the actual number of times the interest each PV is compounded (Benefit I. and Benefit II. period)

Calculation:

The calculation is based on the assumption that weekly planned values would have been accrued until the actual week when the over-planned element would have been released and re-invested therefore the total interest earned on re-investment is calculated based on the difference (PV); the interest in the given scenario (r = 1% or 2% or 3% or 5%), weekly compound (T = 52) and the Benefit I. Period. As from the following week the over planned element would have been re-invested, the only additional benefit is the interest or return earned on the Benefit I interest until the end of the year therefore the Benefit II. calculation is based on the Benefit I.: Interest earned on re-investment (PV); the interest in the given scenario (r = 1% or 2% or 3% or 5%), weekly compound (T = 52) and the Benefit II. Period. Please find the detailed calculation below.

Calendar Period	Benefit I. Interest Period	Benefit II. Interest Period	Difference between Plan and Forecast	Benefit I.: Interest earned on re-investment	Benefit II.: Interest earned on Benefit I. until the end of year	Total Benefit	Cumulative interest when annual interest rate is:			
							1%	2%	3%	5%
206	1	51	£45,700	£8.79	£0.09	£9	£9	£18	£27	£46
207	2	50	£8,605	£3.310	£0.03	£3	£12	£25	£37	£64
208	3	49	£15,068	£8.695	£0.08	£9	£21	£42	£64	£109
209	4	48	£46,780	£35.995	£0.33	£36	£57	£116	£175	£298
210	5	47	£47,636	£45.822	£0.42	£46	£104	£209	£317	£538
211	6	46	£11,929	£13.771	£0.12	£14	£117	£237	£359	£610
212	7	45	-£18,974	£0.000	£0.00	£0	£117	£237	£359	£610
213	8	44	£41,111	£63.290	£0.54	£64	£181	£366	£554	£941
214	9	43	-£11,860	£0.000	£0.00	£0	£181	£366	£554	£941
215	10	42	£7,590	£14.609	£0.12	£15	£196	£396	£599	£1,017
216	11	41	£28,162	£59.631	£0.47	£60	£256	£517	£783	£1,329
217	12	40	£794	£1.835	£0.01	£2	£258	£521	£788	£1,338
218	13	39	£36,528	£91.425	£0.69	£92	£350	£707	£1,069	£1,815
219	14	38	-£31,357	£0.000	£0.00	£0	£350	£707	£1,069	£1,815
220	15	37	£21,045	£60.788	£0.43	£61	£411	£830	£1,256	£2,132
221	16	36	£40,093	£123.542	£0.86	£124	£536	£1,081	£1,636	£2,775
222	17	35	£39,044	£127.841	£0.86	£129	£664	£1,340	£2,028	£3,440
223	18	34	-£11,125	£0.000	£0.00	£0	£664	£1,340	£2,028	£3,440
224	19	33	£934	£3.417	£0.02	£3	£668	£1,347	£2,039	£3,458
225	20	32	£18,151	£69.938	£0.43	£70	£738	£1,489	£2,253	£3,821
226	21	31	£62,528	£253.004	£1.51	£255	£993	£2,002	£3,029	£5,134
227	22	30	-£751	£0.000	£0.00	£0	£993	£2,002	£3,029	£5,134
228	23	29	£8,326	£36.903	£0.21	£37	£1,030	£2,077	£3,142	£5,326
229	24	28	£76,178	£352.368	£1.90	£354	£1,384	£2,791	£4,221	£7,152
230	25	27	£53,027	£255.527	£1.33	£257	£1,641	£3,309	£5,003	£8,475
231	26	26	£12,542	£62.863	£0.32	£63	**£1,704**	**£3,436**	**£5,196**	**£8,800**

A15. The Structure of the Interviews

General introduction

1. Author to describe current research (aim, method, potential applications).
2. Roles and responsibilities of the interviewee at the organisation?
3. Challenges of the organisation in 2016?
4. Challenges of the banking industry in 2016?

Information System

5. What analytical / management information system is operated?
6. What analytical / reporting activity the organisation does conduct?

Planning System

7. What are the key annual marketing objectives in 2016 at your organisation?
8. Please describe the annual marketing planning, especially budgeting process? Probes:
    a. What methods are applied (heuristic and/or analytical)?
    b. What software are used for the planning?
    c. What is the bases of the annual cost budget (previous performance, activity etc.)?
9. Please describe the pros and cons of the existing planning system.

'Brainstorming' session (identification of predictor variables)

10. If Question 7 is heuristic only then probe knowledge on analytical method else skip to 11.
11. Describe the regression approach to cost planning
    a. Brief explanation of mathematical logic behind regression.
    b. Explain what a target variable is.
    c. Explain what a predictor/explanatory/influencer variable/factor is.
12. In your opinion: what are the internal (e.g. operations) and external (e.g. economic) factors that influence/explain/predict the change in the number of new to bank current account, savings and credit card customers? (DRAW!)
13. In your opinion: what are the internal (e.g. operations) and external (e.g. economic) factors that influence/explain/predict the change in the current account, savings and credit card new business market share (campaign market share, not stock market share)? (DRAW!)
14. Other: Do you have any other questions regarding this topic that we have not discussed and you would like to share?

10.14751/SZIE.2017.056

ACKNOWLEDGEMENT

Firstly, I would like to thank my academic supervisor Dr. Zoltán Céman for his continuous support during the doctoral studies. His extended knowledge and experience helped me to approach the research problem and accomplish the goal. I would also like to thank his personal assistant, Krisztina Palotás for the constant support and organisation and the Doctoral School, especially Monika Törökné Hajdú, in Hungary for their professional and also friendly support during these years.

A special thank goes to Zsuzsanna Mészáros for helping me with the methodology. Her statistical and mathematical expertise helped me to identify the correct methods and their pitfalls.

I would like to thank my wife, Szilvia Kalmár for her unconditional support, help with the revisions, patience during the countless hours spent on the research and the encouragement to pursue my interests. Without you it would not have been possible to reach this milestone. Also many thanks to my and her parents Eta Kalmár, János Kalmár, Katalin Ildzsa and Gyula Tóth for all of their support and help.

Last, but not least, I thank my colleagues in the United Kingdom: my manager Mark for shaping my commercial awareness; Fabian and Tim for helping with data collection; and my team Mari, Joseph, Aaron and John for their patience and moral support during the challenging task of balancing between work and current research.