ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Διπλωματική Εργασία

ΔΙΕΡΕΥΝΗΣΗ ΑΠΟΔΟΤΙΚΟΤΗΤΑΣ ΤΥΧΑΙΑΣ ΠΟΛΥΩΝΥΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΕΣΩΤΕΡΙΚΟΥ ΕΛΕΓΧΟΥ: Η ΠΕΡΙΠΤΩΣΗ ΤΩΝ ΕΛΛΗΝΙΚΏΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

του

ΙΑΚΩΒΟΥ Θ. ΜΙΧΑΗΛΙΔΗ

Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού διπλώματος ειδίκευσης στη Διοίκηση Επιχειρήσεων

Απρίλιος, 2019

This page was left blank intentionally.

INTERDEPARTMENTAL MASTER PROGRAM IN BUSINESS ADMINISTRATION (MBA)

Post-Graduate Diploma Thesis

EXPLORATION OF RANDOM POLYNOMIAL REGRESSION EFFICIENCY IN INTERNAL AUDIT SYSTEMS: THE CASE OF GREEK ENTERPRISES

from

IAKOVOS T. MICHAILIDIS

Submitted as a prerequisite for acquiring the post-graduate diploma on Business Administration

April, 2019

This page was left blank intentionally.

Αφιέρωση

Η παρούσα διατριβή είναι αποτέλεσμα δουλειάς μηνών αλλά και επίμονης προσπάθειας, παρόλη την πνευματική και ψυχολογική κούραση λόγω παράλληλων εργασιακών υποχρεώσεων. Έτσι θέλω να την αφιερώσω με όλη μου την καρδιά στον εαυτό μου αλλά και στους ανθρώπους που με στήριξαν: στην σύζυγό μου Κική, στον αδερφό μου Παναγιώτη, στην μητέρα μου Ελένη και τον πατέρα μου Θεόδωρο.

This page was left blank intentionally.

Dedicated to

The current study is a result of monthly work as well as concentrated effort, despite the mental fatigue due to parallel job responsibilities. As a result, I would like to dedicate this study to myself as well as to the ones that supported me: my wife Kiki, my brother Panagiotis, my mother Eleni and my father Theodoros; with all my heart.

This page was left blank intentionally.

Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες στον επιβλέποντα Καθηγητή της παρούσας διπλωματικής εργασίας, Δρογαλά Γεώργιο, Επίκουρο Καθηγητή του Τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Μακεδονίας για την επιστημονική του καθοδήγηση και υποστήριξη σε όλη τη διάρκεια εκπόνησής της. Ταυτόχρονα θα ήθελα να ευχαριστήσω τα δύο άλλα μέλη της εξεταστικής επιτροπής, κύριο Καραγιώργο Θεοφάνη, Καθηγητή του Τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Μακεδονίας και τον κύριο Σταυρόπουλο Αντώνιο, Αναπληρωτή Καθηγητή του Τμήματος Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας για τα πολύτιμα σχόλια τους.

Δεν θα μπορούσα να παραλείψω να ευχαριστήσω επίσης, όλους όσους συμμετείχαν με προθυμία στην έρευνα, οι οποίοι με την άμεση ανταπόκριση τους συνέβαλλαν ουσιαστικά στην ολοκλήρωσή της.

Ακόμη θα ήθελα να ευχαριστήσω και να υπογραμμίσω την αμέριστη κατανόηση και συμπαράσταση του Καθ. Ηλία Β. Κοσματόπουλου με τον οποίο έχω την χαρά να συνεργάζομαι επαγγελματικά τα τελευταία εννέα χρόνια της ζωής μου.

Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα τους γονείς μου Θεόδωρο Μιχαηλίδη και Ελένη Παπανικολάου – Μιχαηλίδη, τον αδερφό μου Παναγιώτη Μιχαηλίδη αλλά και την σύζυγο και συνοδοιπόρο στην ζωή μου κ. Κυριακή Αλεξανδρίδου για την σημαντική ενθάρρυνσή και στήριξή τους, καθ’ όλη την διάρκεια των μεταπτυχιακών μου σπουδών αλλά και της συγγραφής της μεταπτυχιακής διατριβής.

This page was left blank intentionally.

Acknowledgements

I would really like to express my sincere gratitude to the supervisor of the current diploma thesis, Dr. Georgios Drogalas, Lecturer in Business Administration Department of the University of Macedonia for his scientific guidance and support throughout its implementation. In addition, I would like to thank the other two members of the reviewing committee (evaluation board), Pf. Theofanis Karagiorgos, Full Professor in the Business Administration Department of the University of Macedonia as well as As. Pf. Antonis Stavropoulos, Associate Professor in the Applied Informatics Department of the University of Macedonia for their valuable reviewing comments.

I would not miss all auditors and external collaborators that participated actively in the conducted survey which was the main data-generating source for the current thesis.

Moreover, I would like to express my gratitude and underline the inexhaustible understanding and support of Pf. Elias B. Kosmatopoulos, Full Professor in the Electrical and Computer Engineering Department of the Democritus University of Thrace, with whom I have the privilege to professionally collaborate for the last eight years of my life.

Finally, I would like to thank especially my parents Theodoros Michailidis and Eleni Papanicolaou – Michailidis, my brother Panagiotis Michailidis as well as my wife and fellow life-traveller Ms. Kyriaki Alexandridou for their valuable encouraging and support, throughout the entire duration of my post-graduate studies and the current thesis' write-up process.

This page was left blank intentionally.

Περίληψη

Η εταιρική διακυβέρνηση επηρεάζεται ιδιαιτέρως από την υλοποίηση συγκεκριμένων στάνταρτ και την υιοθέτηση διαδικασιών στις σύγχρονες εταιρείες, γεγονός που μετατρέπει τον εταιρικό έλεγχο σε μία κρίσιμη παράμετρο για την βιωσιμότητά τους. Τα εταιρικά σκάνδαλα καταδεικνύουν την ιδιαίτερη ανάγκη για διαφανή και καίρια εταιρική διακυβέρνηση στο σημερινό, ραγδαία μεταβαλλόμενο οικονομικό περιβάλλον. Την ίδια στιγμή, η έλλειψη λειτουργικών κανόνων και μηχανισμών παρακολούθησης επιτρέπει την επέκταση του ήδη υπάρχοντος κενού μεταξύ την σχεδιασμένης και της τελικά υλοποιούμενης λειτουργικής στρατηγικής σε όλους τους τομείς μίας εταιρείας. Σαν αποτέλεσμα τα αυξημένα περιστατικά απάτης σε συνδυασμό με τα μη-διαχειρίσιμα εταιρικά ρίσκα μπορούν να εμποδίσουν την εύρωστη λειτουργία τους. Σε αυτό το εταιρικό και οικονομικό περιβάλλον, ο εσωτερικός έλεγχος αποτελεί καθοριστικό παράγοντα που υποστηρίζει την διοίκηση στην λήψη σωστών αποφάσεων. Ωστόσο η λήψη αποφάσεων σε ένα τόσο πολύπλοκο και ένδο-επηρεαζόμενο οικοσύστημα μπορεί να οδηγήσει στην ανάγκη για επεξεργασία μεγάλου όγκου δεδομένων τα οποία πρέπει να επιλεγούν και να καθαριστούν με προσοχή μέσα από τεράστιες αποθήκες δεδομένων. Για να μπορέσουμε να αντιμετωπίσουμε ένα τόσο επίπονο έργο, χρησιμοποιήθηκαν εδραιωμένες τεχνικές παλινδρόμησης σε ένα σωστά ορισμένο πρόβλημα με εφαρμογή στον εσωτερικό εταιρικό έλεγχο και την αξιολόγηση. Για τον σκοπό αυτό αναπτύχθηκε ένα παραμετροποιημένο σενάριο σε MATLAB σχεδιασμένο ώστε να επιτρέπει πολλαπλές δοκιμές παλινδρόμησης. Ένα γραμμικό ως προς τις παραμέτρους, πολυωνυμικό μοντέλο παλινδρόμησης υιοθετήθηκε για να δοκιμαστούν και να αξιολογηθούν διάφορα πολυωνυμικά μοντέλα Ν-ιοστού βαθμού, χρησιμοποιώντας την κοινή μέθοδο ελαχίστων τετραγώνων σε ένα σύνολο συλλεγμένων δεδομένων από εκτελεστικούς εσωτερικούς ελεγκτές εισηγμένων ελληνικών εταιρειών (τελευταίο τρίμηνο του 2018). Η ανάλυση των συλλεγμένων δεδομένων υπογραμμίζει την ισχυρή αλληλεπίδραση μεταξύ της ποιότητας του εσωτερικού ελέγχου και της εκτελούμενης λειτουργίας εσωτερικού ελέγχου.

Λέξεις Κλειδιά: Εσωτερικός έλεγχος; Εταιρική Διακυβέρνηση; Ελεγκτική συνάρτηση; Ελεγκτική λειτουργία; Εταιρικό μάνατζμεντ.

This page was left blank intentionally.

Abstract

Corporate governance is strongly affected by the implementation of certain standards and followed procedures in modern firms, rendering firm auditing as a crucial parameter for their sustainability. The corporate scandals depict the vigorous importance of transparent and solid corporate governance in todays, rapidly emerging financial environment. At the same time, the lack of operational rules and monitoring mechanisms allows for expanding further the already existing gap between the designed and the actually followed operational strategy in all operational sectors of a company. As a consequence, increased fraudulent incidents accompanied by highly-uncertain business risks may therefore hinder business' robustness. In this business environment, Internal Audit is a key factor in aiding the administration to make right decisions. Decision making for internal auditing, in such a complex inter-affected ecosystem though may suggest for processing a great amount of data which have to be carefully selected and cleansed out of huge data-lakes and warehouses. In order to cope with a cumbersome human task, well established regression techniques are being utilized over a certainly defined problem on internal auditing and control application. A parameterized MATLAB script instance designed so as to enable multiple regression tests was developed for such a purpose. A linear in the parameters, polynomial regression model was utilized in order to test and estimate several different polynomial models of abstract-order using a common least-squares fitting over a surveyed dataset among chief executive auditors from Greek Stock Exchange (listed at the last quarter of 2018) companies. The analysis over the surveyed results highlight the strong interdependence between internal control quality, and internal auditing function execution.

Keywords: Internal audit; Corporate governance; Auditing function; Audit control; Enterprise management.

Table of Contents

Table of Tables

Table of Plots/Diagrams

Chapter 1: Introduction

This page was left blank intentionally.

1.1. Background

Internal auditing topic and self-assessment in organizations has been studied systematically for many decades now. Despite being of core importance for the smooth corporate operational strategy execution, auditors were frequently not perceived as a part of the corrective factors within enterprises especially small-medium ones, which usually did not even follow the basic financial legislation/rules to avoid imposed fees which can be critical for their own viability.

Effective internal control execution is the core ingredient for high-quality financial reporting.

Within the recent past years severe financial and fraud scandals have been reported in the press, suggesting the important need for more effective corporate governance (Endaya & Hanefah, 2016) (Sultana, Singh, & Van der Zahn, 2015) (Khlif & Samaha, 2014). As a result, the significant contribution of internal corporate control function in ensuring financial reporting reliability has been revisited in recent literature (Khlif & Samaha, 2016) (Salehi & Bahrami, 2017) (Bedard & Graham, 2011) (Lin, Wang, Chiou, & Huang, 2014) (Burton, Emett, Simon, & Wood, 2012) (Prawitt, Smith, & Wood, 2009).

As external and internal corporate environments are affected by much more complex dynamics as well as unpredictable disturbances, presenting high stochasticity, the accounting-auditing community is now called to resolve recording problems with increased transaction volumes, scale and complexity. One of the main characteristics, as in all real-life systems, imposed also by the unavoidable natural evolution of fraudulent strategies and tools as well as the diversities in human psychology/behavior, is the constantly changing interplays among different corporal-functioning factors.

Holistic enterprise resource planning (ERP), which synthesizes all operational processes as a whole to achieve centralized performance and profit optimization based on digital records and logs, suggests for a much more complex accounting system. Auditors are called to utilize the capabilities of modern ERPs, which can provide big amounts of recorded data from well-monitored and widely-spread channels (e.g. e-commerce, online payments, etc.), to explicitly or implicitly detect and recognize valuable accounting insights and information. It is auditors' responsibility to effectively audit sufficient amount of recorded data to evidently and unarguably justify their audit opinion, to eliminate unexplored risks-impact-assessment as well as interpret the underlying evidences.

As a result, internal control must be adaptive enough to cope with a dynamic, cumbersome to model and highly stochastic problem within a reasonable amount of response time.

Flexible and updateable internal auditing can guarantee adaptivity and robustness in internal corporate control execution, able to harmonize diverse corporal key performance indicators: value-chain symbiosis, resource use optimization, regulatory and legislation framework compliance, employees' satisfaction, executive governance revenues, accounting transparency, etc. Due to the recent exponential advances in micro-processing sector, more tedious and computationally demanding software tools are steadily penetrating into the auditing procedures to complement the traditional manual audit process.

ERP and corporal governance software auditing tools usually enable data examination through common database query data-requesting or data-visualizing commands which may provide the framework for analyzing and extracting intuitive insights but lack on information mining[1] and patterns profiling/recognition.

On the other side, manual information extraction from recorded data has been of main research interest for the last centuries. Early methods of identifying patterns in data include Bayes' theorem and regression analysis. As computer technology for storing, handling and processing capacity has significantly expanded over the years, automated data analyzing software methodologies and tools have emerged, blossoming upon well-established computer science methods: neural networks, cluster analysis, genetic algorithms, decision trees and decision rules, and support vector machines (Han, Pei, & Micheline, 2011) (Witten, Frank, & Hall, 2011) (Data mining, 2019).

It is more than evident that information-mining can complement and supply auditing with powerful automated tools which can assist auditors' situation analysis and ultimately provide a systematic approach for corporate decision-support and strategic planning. As commented also by the Institute of Internal Auditors: "the support for management in the discharge of these responsibilities is a legitimate role for internal auditors" (IIA, 2004).

1.2. Research Objective and Contribution

The internal audit function/execution (IAF) strategy is usually considered as the guaranteeing factor for implementing high-quality internal control. However, research efforts have not yet merely been focused on the internal control quality effects caused by

[1] *"Information-Mining" term is considered to represent better the exact same scientific topic as the much more widely used buzz-word "Data-Mining". Both terms are equivalently used herein.*

the IAF specifications. Research on internal control weaknesses disclosure to the IAF practices dependence (Fadzil, Haron, & Jantan, 2005) (Lin, Pizzini, Vargus, & Bardhan, 2011), is in general quite scarce in literature. Moreover, they are limited in a twofold manner:

- The analyzed IAF practices usually do not consider the effects of co-working between internal audit and the audit committee, presenting only few limited literature discussions (Barroso-Castro, Villegas-Periñan, & Casillas-Bueno, 2016) (Pugliese, et al., 2009) (De Silva Lokuwaduge & Armstrong, 2015).
- Usually, due to easy data access and high data availability, these studies focus on the US enterprises' case which limits their conclusions "replicability" in other specific countries particular cases (Khlif & Samaha, 2014) (Kinney Jr., Palmrose, & Scholz, 2004) (Hope, Thomas, & Vyas, 2013) (Becker, DeFond, Jiambalvo, & Subramanyam, 1998) .
- Due to the increased application complexity, relevant scientific studies do not attempt to bridge the existing gap between accounting / auditing and information mining topics (Al-Khaddash, Al Nawas, & Ramadan, 2013) (Rezaee, Sharbatoghlie, Elam, & McMickle, 2002).

As a result, research efforts counteracting the aforementioned limitations, in the Greek paradigm, does lack on results and scientific attention. Greece is the southeastern border country of the European continent. Its gross domestic product per capita was $20,311 (see Figure 1), and its unemployment rate was about 19.85% (see Figure 2) based on the collected data since October 2018 (IMF, 2018).

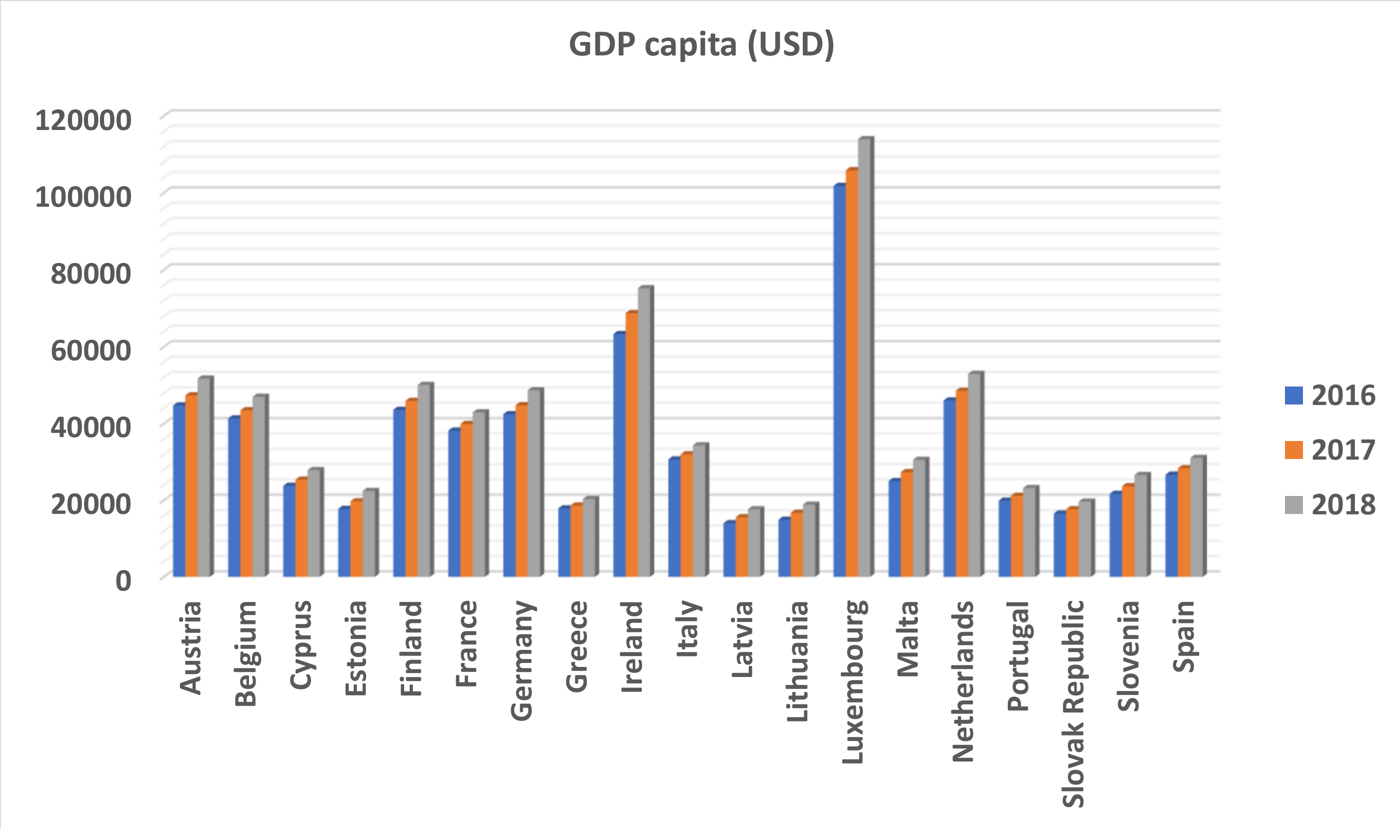

Figure 1. Gross Domestic Product (GDP) per capita in Eurozone (October 2018).

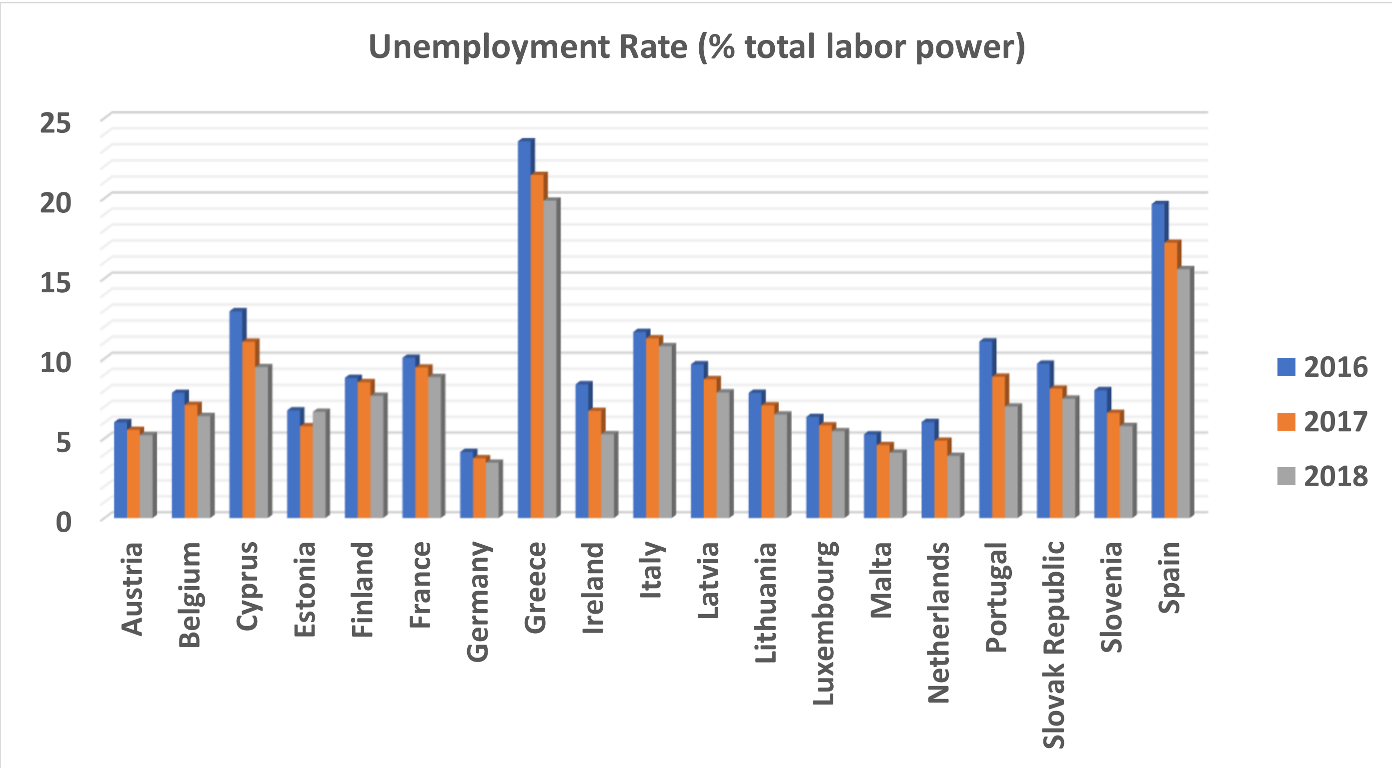

Figure 2. Unemployment Rate as a percentage of the total labor force in Eurozone (October 2018).

Greece is also among the first emerging countries that have thought to legislate on corporate governance (Nerantzidis, Measuring the quality of the “comply or explain” approach, 2015) (Nerantzidis, 2016). In 2001, a regulatory framework for audit committees was enacted for the first time. This law stipulates that credit institutions have to establish an internal control system and an audit committee. However, the Greek regulations still remain limited as compared to the international case. For example, the Sarbanes-Oxley Act of 2002, (2002) requires that management include in its annual report a formal assessment of the effectiveness of its internal controls over financial reporting. Nonetheless, the Greek regulations do not give these parties the possibility of closely monitoring the results of IAFs work (Nerantzidis, Filos, Tsamis, & Agoraki, 2015). They turn into black boxes where doubts might be raised with regard to their effectiveness.

As a conclusion, this thesis focuses on the application of information(data) mining tools on internal auditing cases, examining the impact of the internal audit function (IAF) practices on the corporal internal audit, in the Greek enterprises’ paradigm. The ultimate goal of the current thesis is to investigate the unexplored effect of IAF on internal audit, using different systematic approaches for effectiveness and concrete benchmarked evaluation, based on an adequately large (size = 78) sample from the Greek case.

Based on a sample of 48 out of 78 Greek stock market companies (see section 5.2), it is shown using different techniques, that the internal control quality is significantly and positively dependent (positive correlation factor) with the IAF implementation and audit evaluating board relationship with their co-workers.

1.3. Structure of the Diploma Thesis

The remainder of this thesis is structured as follows:

- Chapter 2, is a brief introduction to auditing. It introduces some essential auditing terms as a basic background. The audit objectives, audit engagement processes and audit approaches are also described here.
- Chapter 3, introduces data mining. Data mining process, tools and techniques are reviewed. Also, the discussions will attempt to explore the concept, methods and appropriate techniques of each type of data mining patterns in greater detail. Additionally, some examples of the most frequently used data mining algorithms will be demonstrated as well.
- Chapter 4, illustrates the actual study. The hypotheses development, relevant facts of the research processes and the study results are presented. Finally, the interpretation of study results will be attempted.

- Chapter 5, describes the selected research methodology, the sample description and the formulated regression models to be used as well.
- Chapter 6, discusses briefly the analyzed results for all direct and indirect measured variables as well as the supervised model training methodology.
- Chapter 7, consolidates briefly the results, the conclusions as well as the proposed future work and steps.
- Chapter 8 (Annex) provides all analytic numbers and values measured and formulated within the selected research methodology.

Finally, the last chapter considers the complete list of relevant references and bibliography of the current study.

Chapter 2: Auditing Framework Overview

This page was left blank intentionally.

2.1. Introduction

The objective of this chapter is to introduce the background information on auditing. A brief historical analysis, definitions of essential terms, main objectives as well as the respective roles of auditing are covered. The auditors' principal roles and engagement stages within the enterprise functioning are also discussed. Audit approaches including test of controls and substantive tests are discussed in greater details while, finally, a brief summary of auditing perspective is provided (Alvin & Loebbecke, 1999).

2.2. Internal Audit History

Internal Auditing is similar to financial auditing by public accounting firms, quality assurance and banking compliance activities (Prawitt, Smith, & Wood, 2009) (Internal audit, 2019). Victor Z. Brink and Lawrence B. Sawyer are considered to be the two "fathers of internal auditing", despite the fact that internal auditing techniques originate from management consulting and public accounting professions (Ramamoorti, 2003).

With the declaration of the Sarbanes-Oxley Act 2002 (section 404) in the United States, internal auditing did establish a more solid position and value. It became evident that internal auditors were the experts having the appropriate qualifications and skills to conform companies with the legislation framework.

2.3. Organizational independence

In order to implement a realistic yet transparent internal auditing function, professional internal auditors must (being also obliged by the IIA standards) be objective despite the policies imposed by the company which actually employs/pays them (Gramling & M., 2006) (Fraser & Henry, 2007) (De Zwaan, Stewart, & Subramaniam, 2011). Independence and objectivity are achieved through the organizational structure layering: for example, in US, internal auditors are mandated to report directly to the board of directors (or a representative audit committee), for management and personnel financial-related activities in order to avoid direct conflict with management and perform an effective and transparent evaluation of the management itself. Internal auditing placement may seem dysfunctional since auditors are paid and managed by the same entity that they are called to objectively evaluate on behalf of the board of directors (or typically the audit committee) (IIA, 2016).

2.4. Role in internal control

Nigel Turnbull developed, in close collaboration the London Stock Exchange for listed companies, a report formalizing the obligations of the directors (Turnbull, 2005) under the Combined UK's Code in order to:

1. Facilitate effective operation by enabling it to respond in an appropriate manner to significant business, operational, financial, compliance and other risks to achieve its objectives.
2. Safeguarding of assets and ensuring that liabilities are identified and managed.
3. Ensure the quality of internal and external reporting, which in turn requires the maintenance of proper records and processes that generate a flow of timely, relevant and reliable information from both internal and external sources.
4. Ensure compliance with applicable laws and regulations and also with internal policies.

The guidelines were also recently revised to comply and meet modern corporate governance requirements.

In principle, internal auditing aims at evaluating the internal control procedures implemented by the company's management.

Finally, internal control is a holistic function/process which comprises five elementary components (Schneider & Becker, 2011), on which managers are called to design and implement processes, policies and practices to ultimately achieve the 4 objectives discussed above:

- the control environment;
- risk assessment;
- risk focused control activities;
- information and communication; and
- monitoring activities.

Consequently, internal auditors are responsible to evaluate the implemented, by the managers, internal control over the five components and ensure (providing standardized recommendations for improvement) its operational effectiveness.

2.5. Role in risk management

To effectively evaluate internal auditing on the company's risk management quality, the respective audit function must be available, according to internationally accepted professional standards. Risk management includes activities which aim at monitoring, detecting, recognizing, analyzing and ultimately mitigating strategic risks that potentially could impact the operational effectiveness of the company.

The most common strategy in conducting risk-driven management implements: (a) estimate/perceive the significance of the risk, (b) assess the severity/impact of the risk scenario, (c) assess the likelihood of the risk and (d) eventually design processes to mitigate the risk (Mueller & Carter, 2007). In particular, the risk scenarios product of significance, severity and probability factors are hierarchically treated accordingly from the highest to the lowest one. These risk mitigating processes provide a roadmap for the foreseen control activities implementation (Morrill, Morrill, & Kopp, 2012).

Risks, and consequently risk management, is directly originating from all vital operations and functions which impose strategic business risks. Management is responsible for assessing risks lurking in ordinary business activities driven by the internal (weaknesses and threats) and external environment (political, environmental, societal, technological) as well.

2.6. Role in corporate governance

The internal auditor is often considered one of the "four pillars" of corporate governance, the other pillars being the Board of Directors, management, and the external auditor. According to COSO-ERM (Steinberg, Everson, Martens, & Nottingham, 2004), corporate governance comprises the policies, processes and structures used by company's leadership to direct activities aiming at fulfilling the company's goals being consistent with ethical standards.

One of the main responsibilities of internal auditors, relevant to corporate governance, is supporting the respective audit committee of the board of directors to execute in compliance with its responsibilities effectively.

2.7. Audit Engagement Processes

Despite auditing execution processes being differently implemented in different auditing companies, auditing in principle involves four major steps (Sirikulvadhana, 2002):

- client continuance: is a process which the auditing firm decides on being (or not) engaged by this client; if the client is accepted, an annual written confirmation (consent), is established and signed. Client acceptance is driven by the following factors and indices:
    - Imposed Engagement Risks: low-ranked clients are in dubious businesses or have too complex financial structure.

  - Relationship Conflicts: to ensure audit quality and to establish a trustworthy interaction with third parties.
  - Clients Requests: including, among others, the qualification of the auditor, working-time constraints, response-time limitation, extra reports and estimated budget.
  - Cost-Benefit Analysis: comparing the potential costs deriving from the engagement with the specific client and the audit fee the client is willing to pay respectively.
  - Collaborative Personnel: Sufficient number of responsive and assistive employees committed to support auditing when and as needed.
- planning: refers to developing an audit plan and includes:
  - team mobilization: refers to forming the auditing team and to communicating among team members.
  - business information elicitation: this step must first establish a smooth cooperation between the client and the audit team in order to collect transparent and raw business data records relevant to the client's industrial sector background, operational structure, implemented control function assessment.
  - risk assessment: the risks assessed are directly linked to the recommended audit strategy, thus, defining a representative formulation of these risks is critical.
  - audit program design: selecting the most suitable audit function, based on the outcomes of the previous steps.
- execution and documentation: involve audit plan execution; results evaluation; observations reporting; summary documentation; results level of acceptance; potential dysfunctionalities and issues detected; recommendations and guidelines for improving the detected issues.
- completion: outputs from the aforementioned steps are concentrated, recorded and formalized to be ultimately reported back to the client. The chief-executive-auditor (CEA) is then responsible for the final review quality upon delivery.

2.8. Auditing Practice

To detect financial statement frauds auditors usually perform tests to extract sufficient evidence. The number and type of audit tests depend completely upon auditors' professional experience and judgement. Audit tests can be categorized in two main categories:

- **Tests of Controls:** Testing the effectiveness of audit control during the specified auditing period. Only key audit control functions will be tested while the number and type of tests depends solely on the control reliance level. The higher control reliance is, the more tests are performed.
- **Substantive Tests:** Substantive test is used for assessing the integrity and transparency of the financial statement balances, in cases when the tests of controls could not be performed due to low control reliance or poor/insufficient evidence.

This page was left blank intentionally.

Chapter 3: Data Mining Framework Overview

This page was left blank intentionally.

3.1. Introduction

The objective of this chapter is to describe the basic concept of data mining and the appropriate background on data mining. Moreover, this chapter is an attempt to describe in a more detailed manner the most popular data-mining processes, tools and techniques.

3.2. Data Mining Domain

Despite the fact that the “data mining” term is usually used to represent a large group of cognitive techniques for behavioral or pattern identification, it is common in literature to replace it with “information mining” or “information extraction” or “knowledge mining” in order to denote the data-driven learning features of these techniques. Data mining (or information mining) techniques were initially studied to support learning and emulating behaviors/reactions of unknown (considered as a black box) natural or technical systems of specific interest (Hand, Mannila, & Smyth, 2001) (Han, Pei, & Micheline, 2011).

Within recent past years, due to the technological advances on micro-processing and integrated large-scale chips, devices with very small size, quite cheap to purchase, easy to install, easy to connect, usually light and mobile (battery powered) emerged rapidly, creating a vast and almost unbounded variety of diverse collected sets of data. Moreover, local and cloud data storing platforms also became quite cheap (if not free in some cases) allowing easy and seamless access to these data from common electronic devices through secure and authorized remote end-points.

To exploit and explore this vast landscape of data, since manual data analysis is mostly impossible, automated data mining techniques did attract scientific and engineering attention, as it promises to extract valuable information for increasing operational efficiency, which can refer to analytically formulated key performance indexes (KPIs), depending each time on the type of the application.

These techniques are designed to automatically manipulate and utilize big amounts of structured data to uncover/discover new, hidden or unexpected trends, patterns, information or behavior. Ultimately, the revealed patterns will allow for future system behavior prediction, off course with certain expected levels of probable error, which will equip the system operator or stakeholder to optimize his operations in a pro-active manner.

3.3. Data Mining Categories

Tailored data mining techniques for different kinds of structured data (e.g. relational database, transactional database, object-oriented database, etc.) and data types (e.g.

spatial database, time-series database, text or multimedia database, legacy database and the World Wide Web) have been proposed in literature (Chapman, et al., 2000).

It must be underlined that the terms Data and Big-Data Mining are used to indicate their capability to cope with large data sets but do not mean that the complete data set involving huge amount of data is required as input. In fact, it is not the quantity rather than the quality of data that is considered much more important in all cases. Aside from containing representative behaviors – as much as possible - for the whole population, the data should reflect normal steady-state or transmission-period behavior of the subject-system with low (or zero) noise (i.e. false, abnormal, stochastic behavior, etc.) to eliminate respective noisy effects on information mining.

The goal of data-mining can be categorized into two groups to predict future trends and behaviors (see Figure 3) (Zhu & Lin, 2007):

- situation verification (top-down approach): Data-mining is utilized to validate and confirm selected hypotheses or to understand/systematize observed situations. To test such hypotheses, certain solid data need to be carefully selected and analyzed. Eventually, after the analysis, an extrapolating object/entity will enable future behavior estimation or clustering. The disadvantage of this approach is that the considered hypotheses or situations are driven by the experience of the analyst.
- pattern discovery (bottom-up approach): Data-mining process is built to automatize the data landscape exploration for recognizing any underlying unknown, complex or unordinary patterns. To uncover such patterns, almost the complete data set is usually required to detect and recognize any interesting underlying data/variable associations. However, analysts are again required to decide upon the value of the mined results.

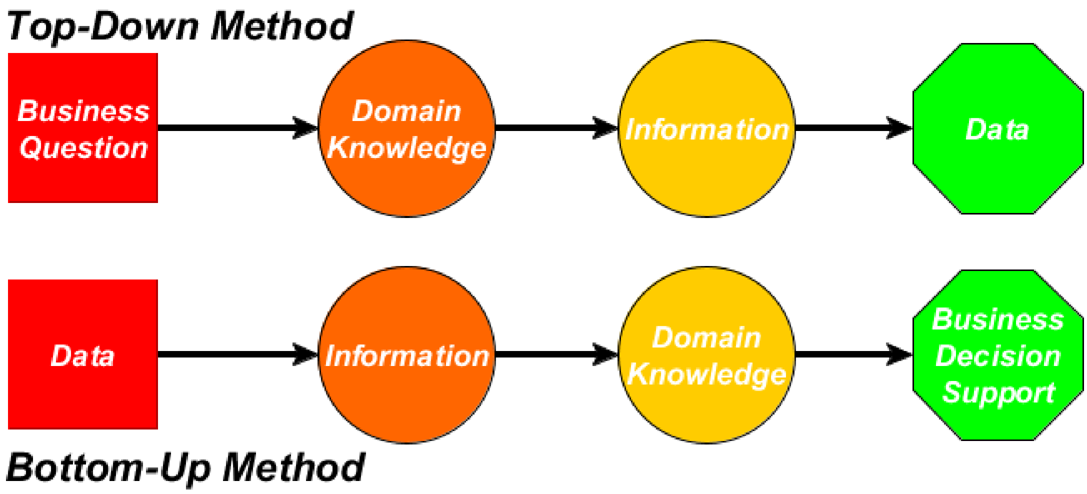

Figure 3. Data Mining Categories.

Both approaches are complementary and not overlapping, since the top-down approach results data and facts which are the starting point of the bottom-up method to eventually result business decision and conclusions extraction in a cyclical fashion.

3.4. Standardized Data Mining Process Layers

Cross-industry standard process for data mining (CRISP-DM - (Chapman, et al., 2000)) is an open standard process model that describes common approaches used by data mining experts. CRISP-DM standardized data mining into a layered process with four levels (see Figure 4) (Shearer, 2000) (Sheikh, 2011).

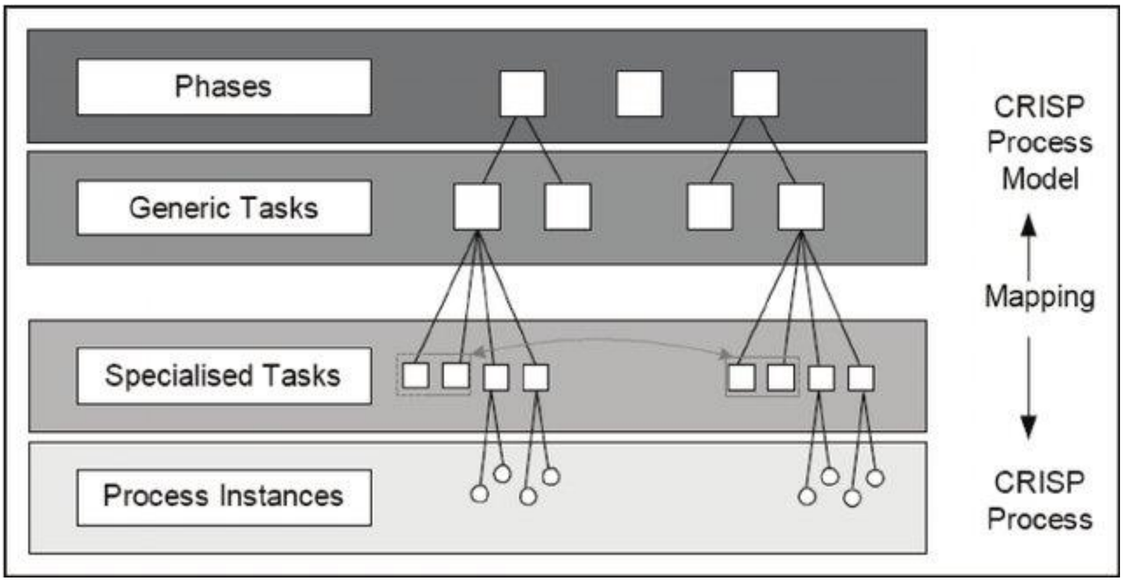

Figure 4. Process Model from CRISP-DM-2000-W640 (http://www.crisp-dm.org/).

- Layer 1 - Phases: corresponds to the data mining processes for deploying mining methods which try to solve certain problems.

- Layer 2 – Generic Tasks: Each phase from Layer 1 contains more specialized tasks for dealing and coping with certain situations.
- Layer 3 – Specialized Tasks: To elaborate more on the level of details, the generic tasks from Layer 2 are greater detailed. At this stage, questions like: e.g. how, when, where and who, are formally answered in order to design the process execution plan.
- Layer 4 – Process Instances: Finally, this layer, corresponds to recording all actions, decisions and results taken out of each phase foreseen in Layer 1.

3.5. Standardized Data Mining Phases

The data mining processes can be algorithmically analyzed into six phases (Layer 1 of CRISP- Data Mining Process – see Figure 4) as follows, shown also in Figure 5:

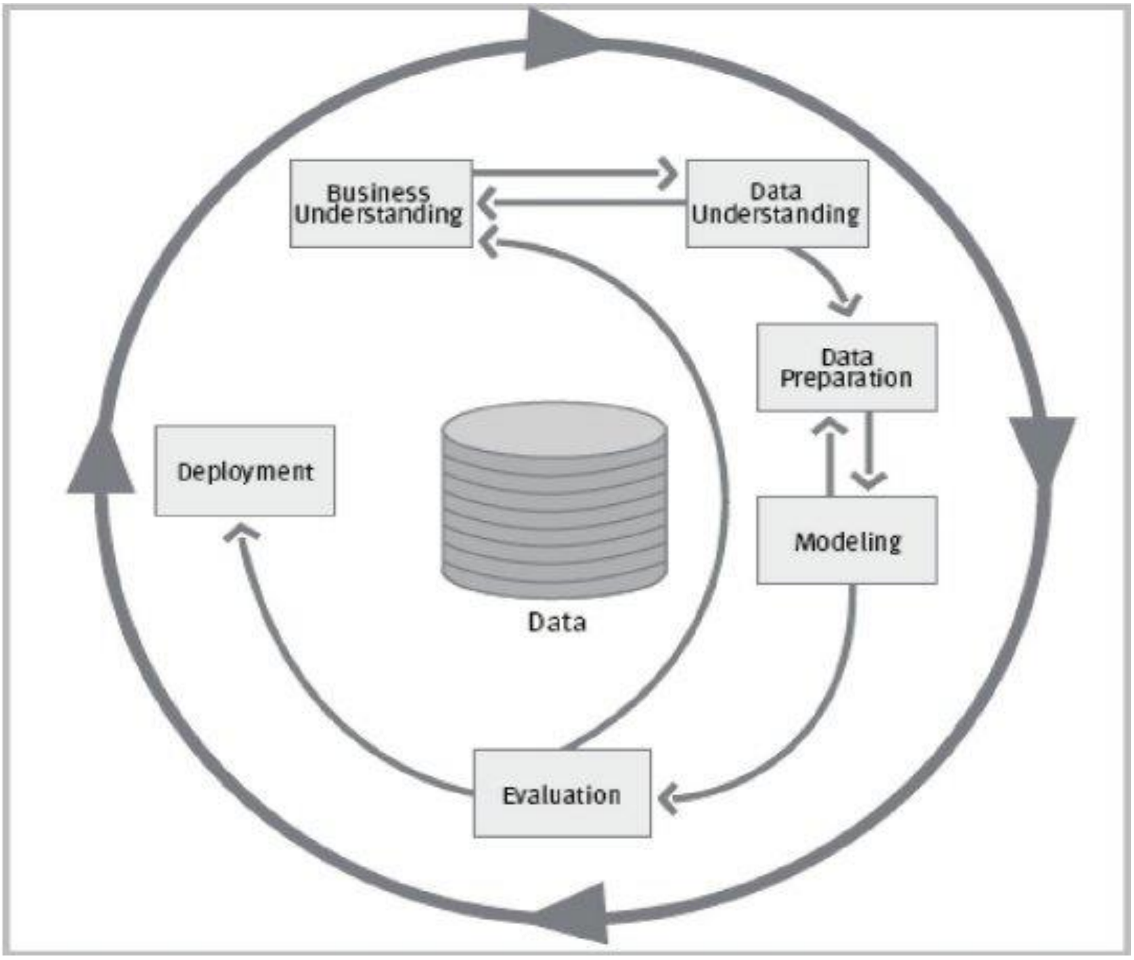

Figure 5. CRISP-DM-2000-W640 Phases (http://www.crisp-dm.org/).

- business understanding: mapping of business issues and objectives to produce the implementation plan as a guideline to the whole project.
- data understanding: includes collecting, describing and defining the type and physical/natural/expected correlation of certain available data sets.

- data preparation: ensure that the data sets are ready to be mined i.e. deciding on which data is relevant, inconsistent data cleaning, data (recreation) aggregation integration and structuring, data transformation into standardized format.
- modeling: build a model that most satisfactorily describes/emulates the global data population set: methodology selection, model design, model training, model testing (in and out of the predefined training set), model validation and model assessment (quantified expected model error).
- evaluation: the obtained results have to be evaluated according to the identified business objectives. At this point, if the evaluated model is accurate enough the process moves to the next phase, otherwise, another training/modeling iteration or even moving back to the very first phase is needed to reengineer the process from scratch.
- deployment: results are deployed and adopted into the business operations and implementation procedures in this phase.

It has to be underlined at this point that such model's efficiency and capabilities are highly dependent on the available data, the recreated/aggregated data, the model design as well as on the selected training data and the methodology used (Berson & Thearling, 1999) (Rud, 2001). Note that its ability to efficiently operate into a very wide range of occasions and situations which are considered definite to occur during the course of the business life-time is quite ambiguous. Model-adaptivity important though since decision making is necessary also during observed business (or exogenous) uncharted behavioral divergences. Usually to cope with such a problem, automated fine-tuning techniques or manual model re-evaluation (or reconstruction) is performed at a regular basis.

3.6. Data Mining Related Approaches Overview

In principle data mining tools are designed for coping with huge amounts of data in contrast to traditional statistical tools. However, data mining is a subdomain of statistics (mathematics) which has been alternated and refined to incorporate more specific functionalities/features according to each application needs over the years. Therefore, data mining approaches may achieve similar (or even the same) goals but may differ in the approach used which is directly linked to the sector for which it was originally developed for. The approaches can be categorized into four broad categories:

- database tools: data generalization, data normalization, missing data detection and restoration, data aggregation, data transformation, attribute-oriented induction, and fractal and online analytical processing (OLAP).

- statistical tools: mean, median, variance, standard deviation, probability, confident interval, correlation coefficient, non-linear regression, chi-square, Bayesian theorem and Fourier transforms.
- artificial intelligence tools: neural network, pattern recognition, rule discovery, machine learning, case-based reasoning, intelligent agents, decision tree induction, fuzzy logic, genetic algorithm, brute force algorithm and expert system.
- visualization tools: for multidimensional data sets in various formats (more elaborate presentation techniques) that allow users to manually (human effort to analyze and assess the results) explore complex multi-dimensional data, e.g. audio, tabular, scatter-plot matrices, clustered and stacked chart, 3-D charts, hierarchical projection, graph-based techniques and dynamic presentation.

3.7. Data Mining Methodologies Overview

Despite the fact that data mining has mostly been automatized utilizing the programming capabilities of even more computationally intense algorithms in continuously empowered computing platforms, the user is still required to determine the methodology which should be applied to the problem under consideration. In order to do so data analyzers and tools users should have the basic features that discretize these methods' applicability. The most common categorization of data mining methodologies is as follows:

- **Data Description:** to generate an overall summarized, consistent description of data itself or in a classified manner either through data characterization i.e. summarizing general dataset characteristics or through data discrimination/comparison i.e. comparison of data characteristics belonging to contrasting groups or classes.

  Often data description helps to gain insight into the data nature or to compactly present classified data features. Appropriate data mining techniques usually used for this purpose are: attribute-oriented induction, data generalization and aggregation, relevance analysis, distance analysis, rule induction and conceptual clustering.
- **Dependency Analysis:** to explore the most statistically significant correlation across huge datasets including large numbers of variables and/or features. A common application is the "basket analysis", analyzing with certain confidence percentage which products customers choose to buy together. Appropriate data mining techniques usually used for this purpose are: nonlinear regression, rule

induction, statistical sampling, data normalization, Apriori algorithm, Bayesian networks and data visualization.

- **Classification and Prediction:** to build models/classifiers which are functions actually that link/group data records to the most relevant class of the already prescribed ones. Model construction requires two data sets, the training/learning one i.e. for training and building an as much as accurate as possible classifier, and the testing/validating one i.e. which will test the built model's capabilities to automatically categorize and classify unknown data records to the existing classes. If the results of testing are unsatisfactory, then more training iterations are required. Appropriate data mining techniques usually used for this purpose are: neural networks, relevance analysis, discriminant analysis, rule induction, decision tree, case-based reasoning, genetic algorithms, linear and non-linear regression, and Bayesian classification.
- **Cluster analysis**: to filter/separate/purify data with similar characteristics from the rest. Despite being usually confused with classification, clustering filters data on-the-fly without utilizing any pre-trained model (unsupervised learning) in contrast to classification which does (supervised learning). Appropriate data mining techniques usually used for this purpose are: neural networks, data partitioning, discriminant analysis and data visualization.
- **Outlier Analysis:** outliers (data records which are dissimilar to others) are usually treated as noise/error data which should be filtered prior to feeding them into a purified-learning model-training process. Surprisingly though, noise presence may occasionally hold useful information for unveiling abnormal behaviors/patterns. Therefore, useful outlying data (outliers) must be filtered-out from noisy/faulty behavior caused by errors and malfunctioning. Filtering may be substituted by finding and fixing erroneous data instead of simply neglecting them. Appropriate data mining techniques usually used for this purpose are: data cube, discriminant analysis, rule induction, deviation analysis and non-linear regression.
- **Evolution Analysis**: to detect, recognize and ultimately predict the underlying trends (noise and mixed harmonics) of certain data-sets throughout the evolution of time in large data warehouses i.e. evolution analysis refers to the extended versions of other methodology types with time-analysis capabilities. Appropriate data mining techniques usually used for this purpose are: sequential pattern discovery, time-dependent analysis.

3.7. Data Mining and Data Analysts

Data mining methodologies may usually spot and reveal unknown patterns and trends in data, however not all mined information is valuable or interesting. To exploit such pieces of information they must be easily understood, valid and potentially useful (Han, Pei, & Micheline, 2011). Therefore, analysts are still needed in order to evaluate whether the mining results are interesting.

To filter valuable information from non-relevant or non-practical observations data mining tools must be capable of:

- **Measuring the correctness of patterns:** based on existing historical data sets, currently observed ones are compared in parallel to classify them accordingly.
- **Creating the optimization model of mined patterns.**
- **Systematically define the correct termination point:** the accuracy level of the currently optimized pattern and the utilization process of such generalizing pattern on other unknown, possibly exogenous, data sets (over-fitting problem).

Chapter 4: Internal Audit Control and Effectiveness

This page was left blank intentionally.

4.1. Introduction

The effectiveness of the Internal Auditing Function is closely connected with the corporate governance environment; however, this relationship has never been quantified in an analytic manner using deterministic and statistical tools (Lenz, Sarens, & D'Silva, 2014). Depending on the composite measure (index) that is considered as representative for the IAF effectiveness, several diverse studies have been already published in literature (Alzeban & Sawan, 2015) (Soh & Martinov-Bennie, 2011) (Sarens, De Beelde, & Everaert, Internal audit: a comfort provider to the audit committee, 2009) (Alhajeri, 2017) (Prawitt, Smith, & Wood, 2009) (Johl, Subramaniam, & Cooper, 2013) (Pizzini, Lin, Vargus, & Ziegenfuss, 2015).

This section of the current study, similar to the above, is an attempt to set up and formulate the problem of quantifying the impact of IAF characteristics on internal control quality. The theoretical framework considers IAF as a critical reporting mechanism for effective corporate governance aiming at reducing situation awareness inadequacies (Sarens & Abdolmohammadi, 2011) (Goodwin-Stewart & Kent, 2006).

4.2. Hypotheses Development

Studies showing the strong positive relationship between the IAF attributes and the quality of the internal control system addressed in a systematic manner are quite scarce in literature. Within the current study, the association between internal control quality and key IAF attributes is examined based on past studies' findings (Johl, Subramaniam, & Cooper, 2013) (Pizzini, Lin, Vargus, & Ziegenfuss, 2015) (IIA, 2016):

a) ***IAF organizational status,***
b) ***working relationship between the internal auditor and the audit committee,***
c) ***internal audit staff competence,***
d) ***IAF investment,***
e) ***quality assurance program and;***
f) ***the follow-up on internal control deficiencies.***

This section presents the statistical hypotheses-of-study related to the association between IAF attributes and internal control quality (ICQ) as derived by the aforementioned attributes. The framework of the considered analysis problem originates from a recently published work (Oussii & Taktak, 2018).

4.2.1. Internal audit function organizational status

Internal audit activity must be independent, and internal auditors must be objective in performing their work (IIA, 2016). Previous research shows that IAF organizational status is one of the most significant factors positively affecting the dependency of external auditors' decision-making on internal control programs (Bame-Aldred, Brandon, Messier, Rittenberg, & Stefaniak, 2013) (Lin, Pizzini, Vargus, & Bardhan, 2011). Based on the above, a hypothesis can be formed in the same respect as follows:

H1. There is a positive association between IAF organizational status and ICQ.

4.2.2. The relationship between the chief audit executives and the committee

Prior research argued that an effective working relationship between the audit committee and the implemented IAF can (positively affect) establish a more concrete and objective internal auditors' judgement opinion, implying its improved execution/implementation (Alzeban & Sawan, 2015) (Arena & Azzone, 2009). As a result, audit committee's involvement in reviewing the IAF execution program is expected to positively affect ICQ. Thus, the following hypothesis is considered:

H2. There is a positive association between the audit committee's involvement in reviewing the IAF execution and ICQ.

4.2.3. Internal audit function competence

Internal auditors must meet the required skills and other competencies to achieve their goals. Similarly, the internal audit function program must exploit auditors' knowledge, skills, and other competencies to fulfil its mission.

The level of adequacy of the IAF execution, as evaluated by external auditors, should consider the technical competence of the internal auditors (IIA, 2016, p. 6) (Arena & Azzone, 2009) (Mihret, James, & Mula, 2010) (Soh & Martinov-Bennie, 2011).

Technical competence of internal auditors is expected to improve IAF execution effectiveness by consequently eliminating internal control deficiencies. Therefore, the following hypothesis is considered:

H3. There is a positive association between IAF competence and ICQ.

4.2.4. Internal audit function investment

Numerous published studies, converge to the same thesis which concludes that allocating greater resources for IAF could result in higher ICQ with better-skilled internal auditors and far better risk assessment and mitigation mechanisms implemented (Alhajeri, 2017) (Gramling & M., 2006) (Bedard & Graham, 2011).

Therefore, IAF execution effectiveness is expected to have a positive effect to the available tools and resources to the internal audit staff. As a result, the following hypothesis is proposed:

H4. Allocating greater resources for IAF leads to less severe internal control weaknesses.

4.2.5. Quality assurance and improvement program

Besides the recommended continuous monitoring and frequent external independent assessment, one of the main responsibilities of the chief audit executives (CAEs) is to develop and maintain the ICQ assurance program that covers all aspects of the internal audit activity i.e. ethics, standards (IIA, 2016, p. 7). Several literature studies indicate a positive relationship between the quality assurance techniques and ICQ reporting (Lin, Pizzini, Vargus, & Bardhan, 2011) (Pizzini, Lin, Vargus, & Ziegenfuss, 2015) (Johl, Subramaniam, & Cooper, 2013).

These arguments suggest that the use of quality assurance practices is positively associated with internal control quality. Thus, we test the following hypothesis:

H5. There is a positive association between internal audit quality assurance and ICQ.

4.2.6. Follow-up on internal control deficiencies

To achieve a coherent and sound ICQ management, knowledge acquired from past events and abnormal deficient situations can be treated as a concrete basis for designing and adapting the IAF execution by the responsible chief executives. Such a strategy will allow for a follow-up process to monitor and recognize past-observed internal control deficiencies early enough (Lin, Wang, Chiou, & Huang, 2014).

Therefore, the availability of such strategies and procedures is positively associated with ICQ can form the following hypothesis of study:

H6. There is a positive association between the existence of follow-up process and internal control quality.

This page was left blank intentionally.

Chapter 5: Research Methodology

This page was left blank intentionally.

5.1. Introduction

The test population and the research procedure followed for the data and correlation analysis of the IAF and ICQ attributes, described in section 4, are discussed within this section. The empirical raw data collection, preprocessing and structuring is also provided herein in an attempt to pave the road for extracting useful conclusions and information on the same matter.

5.2. Sample – Population

To collect, study and analyze the attributes and hypotheses associating IAF execution and ICQ described in section 4, a survey method was adopted considering a distribution of a targeted questionnaire among chief auditing executives from companies listed on the Greek Stock Exchange during the last quarter of 2018.

In total, 78 questionnaires were successfully distributed electronically, while 48 (61.5% completion rate) exploitable responses were received from the 78 companies, 2 out of 78 responses were inadequately completed and therefore neglected and the 28 remaining invitations were never answered. The size of the usable sample i.e. 48, represents the 26% of all 185 listed companies on the Greek Stock Exchange (http://www.helex.com/el/companies-map). The sample size can be arguably considered large enough to allow for a more representative and concise conclusion extraction for the overall population. Table 1 provides a compact description of the sample studied herein.

Table 1. Sample profiling and breakdown

	PERCENTAGE W.R.T. THE TOTAL POPULATION	**ABSOLUTE NUMBER**
TOTAL NUMBER OF FIRMS LISTED ON GSE AT 12/2018	100%	185
EXCLUSIONS	57.8%	107
FIRMS DELETED BECAUSE OF NON-RESPONSES TO THE SURVEY	15.1%	28
FIRMS DELETED BECAUSE OF MISSING DATA	1.1%	2
FINAL USEABLE SAMPLE	26%	48
SAMPLE ENTERPRIZES ANALYSIS PER SECTOR		
SECTOR	TOTAL NUMBER	INCLUDED/REPLIED

CONSTRUCTION	10	9
STEEL	6	2
FURNITURE	4	3
CLOTHING & ACCESSORIES	10	7
PERSONAL CARE	2	1
HEALTHCARE	5	2
FOCUSED RETAILING	6	5
TOURISM AND TRAVEL	6	4
CONVENTIONAL ELECTRIVITY	2	2
COMPUTER SERVICES	9	8
ICT SOFTWARE	3	3
COMPUTER HARDWARE	3	1
ALUMINIUM	1	1
TOTAL	67	48

5.3. Variable Measurement

5.3.1. The dependent variable

Based on previous studies (Bedard & Graham, 2011) (Oussii & Taktak, 2018) the dependent variable (response of the modelled system), denoted as ICQ, is defined as:

the number of internal control deficiencies detected annually by chief executive auditors,

as a representative index for the quality of internal control. The ICQ values became available from a respective **single item of the distributed questionnaire**. IAF characteristics are expected to elaborate on the internal audit quality and consequently reduce the occurrence of ICQ deficiencies and weaknesses.

5.3.2. The independent variables

As already discussed in the introduction of section 4, six independent variables are used for the considered statistical and relational analysis:

a) IAF organizational status: denoted as IAF_OS, is a **dummy post-designed** variable to test H1, while, as described in section 4.2.1. The variable is designed so as to take the value of one only in case when the IAF reports functionally to the audit committee:

$$IAF_{OS} = \begin{cases} 1, & IAF\ reports\ to\ CAE\ committee \\ 0, & otherwise \end{cases}$$

b) ***working relationship between the internal auditor and the audit committee:*** denoted as WKREL, is a variable used to test H2 as described in section 4.2.2. The assessment of this variable, denoting whether the auditing committee reviews internal IAF program executed by internal auditors, is derived by implementing a five-point Likert scale **questionnaire item**, as follows:

$$WKREL = \begin{cases} 1, & strongly\ disagree \\ 2, & disagree \\ 3, & neutral \\ 4, & agree \\ 5, & strongly\ agree \end{cases}$$

c) ***internal audit staff competence:*** denoted as STF_COMP, is a variable used to test H3 as described in section 4.2.3, obtained by standardized and averaged fusion of **five questionnaire items as follows:**

$$STF_{COMP} = STF_{COMP,P} + min\left(STF_{COMP,P}\right)$$

Where:

$$STF_{COMP,P} = \frac{Z_{EXP} + Z_{EDU} + Z_{CERT} + Z_{TRN}}{4}$$

- experience (denoted with EXP): company’s internal auditors’ average experience standardized number of years, available as a **single questionnaire item**;

$$Z_{EXP} = \frac{EXP - \mu_{EXP}}{\sigma_{EXP}}$$

- education (denoted with EDU): company’s internal auditors’ average higher education standardized number of years (after high school), available as a **single questionnaire item**;

$$Z_{EDU} = \frac{EDU - \mu_{EDU}}{\sigma_{EDU}}$$

- certification (denoted with CERT): the standardized fraction of the number of company’s internal auditors who are certified with at least one audit certification (available as a **single questionnaire item)** over the total number of internal auditors of the company, also available as a **single questionnaire item**;

$$Z_{CERT} = \frac{CERT - \mu_{CERT}}{\sigma_{CERT}}, CERT = \frac{NoCertified}{NoTotal}$$

- training (denoted with TRN): company’s internal auditors’ average standardized number of training hours per year, available as a **single questionnaire item**.

$$Z_{TRN} = \frac{TRN - \mu_{TRN}}{\sigma_{TRN}}$$

d) ***IAF investment:*** denoted as IAF_INV, is a variable used to test H4 as described in section 4.2.4. It has been formulated as the fraction of **a single questionnaire item** (Zain, Subramaniam, & Stewart, 2006)**:** the natural logarithm of the total number of human resources (denoted as IAF_HMR) participating in the IAF execution:

$$IAF_{INV} = ln(IAF_{HMR})$$

e) ***quality assurance program:*** denoted as QAS, is a variable to test H5 as described in section 4.2.5. It has been formulated as a single composite obtained by linearly and evenly (using the sample global averages) fusing **six questionnaire items as follows** (Johl, Subramaniam, & Cooper, 2013):

$$QAS = QAPX + NT\_ASS + XT\_ASS$$

- QA program existence (denoted with QAPX): scalar Boolean variable, available as a **single questionnaire item**, indicating if (value 1) a quality assurance program/plan is being formally implemented or not (value 0);

$$QAPX = H\left(QAPX_P - \widetilde{QAPX_P}\right)$$

- Internal assessment (denoted with NT_ASS): measured as the normalized average of **two** five-point Likert scale **questionnaire items** (from: 1 – none at all; Up to: 5 - always):
  - the utilization of internal continuous monitoring tools (UT_MON);
  - the reporting tendency of periodic auditing reviews (R_TEND).

$$NT_{ASS} = H\left(NT_{ASS,P} - \widetilde{NT_{ASS,P}}\right)\left(\frac{UT_{MON} + R_{TEND}}{\max\,(UT_{MON} + R_{TEND})}\right)$$

- External assessment (denoted with XT_ASS): formulated as the average of **three questionnaire items**:

$$XT_{ASS} = H\left(XT_{ASS,P} - \widetilde{XT_{ASS,P}}\right)\left(\frac{EXQA + EXASS + PREX}{3}\right)$$

  - Denoted as EXQA, the existence of an external quality assessment (Yes=1/No=0);
  - Denoted as EXASS, the implementation of a fully external assessment or self-assessment assisted by external validation (Yes=1/No=0), and;
  - Denoted as PREX, the periodic implementation of an internal auditing external evaluation every five years (Yes=1/No=0).

Note that for all aforementioned cases, the tilde symbol is used for denoting the median value of the respective questionnaire item sampled while $H(x_0)$ function denotes a diversified version of the identity Heaviside step function formulated as follows:

$$H(x - x_0) = \begin{cases} 1, & x > x_0 \\ 0, & x \leq x_0 \end{cases}$$

f) ***the follow-up on internal control deficiencies:*** denoted as FUP_DEF, is a **post-designed dummy** variable to test H6 as described in section 4.2.6. The variable is designed to take the value of one only if the IAF builds upon the knowledge acquired from previously observed internal control deficiencies:

$$FUP_{DEF} = \begin{cases} 1, & IAF\ mitigates\ observed\ with\ past\ deficiencies \\ 0, & otherwise \end{cases}$$

5.3.3. The control variables

Moreover, a number of firm attributes and features are considered as independent control variables, associated with internal control quality (Khlif & Samaha, 2016) (Bedard & Graham, 2011) (Lin, Pizzini, Vargus, & Bardhan, 2011). The 4 control variables used in the model, deriving from respectively designed **single questionnaire items**, are:

- the percentage of financial experts in the audit committee (FC_XP);
- natural logarithm of the entity's sales size (LN_SLS);
- return on assets ratio financial index (ROA) and;
- boolean variable that equals one only if the firm belongs to any financial industry sector (FIN_IND).

5.4. Statistical Regression Methodology

5.4.1. Polynomial 1st Order (Linear) Formulation

In order to test the validity of the designed problem, as already discussed in section 4 and in subsection 5.4 above, a linear statistical regression model is utilized as follows:

$$ICQ \approx \widetilde{ICQ} = \beta_0 + \beta_1 IAF_{OS} + \beta_2 WKREL + \beta_3 STF_{COMP} + \beta_4 IAF_{INV} + \beta_5 QAS + \beta_6 FUP_{DEF} + \beta_7 FC_{XP} + \beta_8 LN_{SLS} + \beta_9 ROA + \beta_{10} FIN_{IND}$$

The model employs a reduced version of a first order polynomial regression where the unknown beta coefficients are scalar real values. The regression analysis strategies utilized in section 6 try to descriptively or analytically calculate/estimate their values using the surveyed dataset sample.

5.4.2 Polynomial N-th Order Formulation

In order to test the validity of the designed problem, as already discussed in section 4 and in subsection 5.4 above, a polynomial statistical regression model is utilized in the general case, as follows:

$$ICQ \approx \widetilde{ICQ} = \beta_0 + \sum_{i=1}^{M} \beta_i \begin{pmatrix} (IAF_{OS})^{n1}(WKREL)^{n2}(STF_{COMP})^{n3}(IAF_{INV})^{n4}(QAS)^{n5} \\ (FUP_{DEF})^{n6}(FC_{XP})^{n7}(LN_{SLS})^{n8}(ROA)^{n9}(FIN_{IND})^{n10} \end{pmatrix}$$

Where "n1, n2, …, n10" are positive integer numbers, denoting the order of each independent variable contributing to the total order of the monomial as follows:

$$n1 + n2 + n3 + n4 + n5 + n6 + n7 + n8 + n9 + n10 \leq N$$

The model employs a version of a reduced N-th order polynomial regression where M is the number of monomials, and as a consequence the number of the unknown beta coefficients which are real scalar values. The linear regression model can be rewritten in the general compact form as follows:

$$\widetilde{ICQ} = \theta \cdot \varphi$$

$$\theta = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_M]$$

$$\varphi_i = \begin{matrix} (IAF_{OS})^{n1,i}(WKREL)^{n2,i}(STF_{COMP})^{n3,i}(IAF_{INV})^{n4,i}(QAS)^{n5,i} \\ (FUP_{DEF})^{n6,i}(FC_{XP})^{n7,i}(LN_{SLS})^{n8,i}(ROA)^{n9,i}(FIN_{IND})^{n10,i} \end{matrix}$$

$$\varphi = \begin{bmatrix} 1 \\ \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_M \end{bmatrix}$$

The regression analysis strategies utilized in section 6.3 try to descriptively or analytically calculate/estimate their values using the surveyed dataset sample.

Chapter 6: Research Results and Evaluation

This page was left blank intentionally.

6.1. Introduction

After the distribution of questionnaires, the data described in section 5 were collected, filtered and structured accordingly. A MATLAB/Simulink script instance was developed to help automatize such a task. Moreover, based on the described statistical methodology, in section 5.5, several tests have been executed adopting polynomial regression models of different maximum order and number of monomials. The statistical analysis results are described within the current section.

6.2. Descriptive Statistics

Using the exact same order of the questionnaire variables as they appear in section 5, indicative descriptive statistical attributes are listed for each variable respectively below. As already discussed in the introduction of section 4, 6 independent variables and 4 control variables are used for the considered statistical and relational analysis for the single scalar dependent variable.

6.2.1. The independent variables

As depicted in Figure 7, over 77% of the collected usable questionnaires consider a functional reporting within the implemented IAF to the internal auditing committee (IAF_OS). Moreover, as shown in Table 2, the regression coefficient (R correlation) of this variable takes a positive value, denoting that **H1** hypothesis (See section 4.2.1) is **true** with a confidence level much smaller than 95% since the p-value = 0.99>0.05 corresponds to a non-significant correlation in R = 0.0012 and a high probability of observing the respective null hypothesis (see Figure 6).

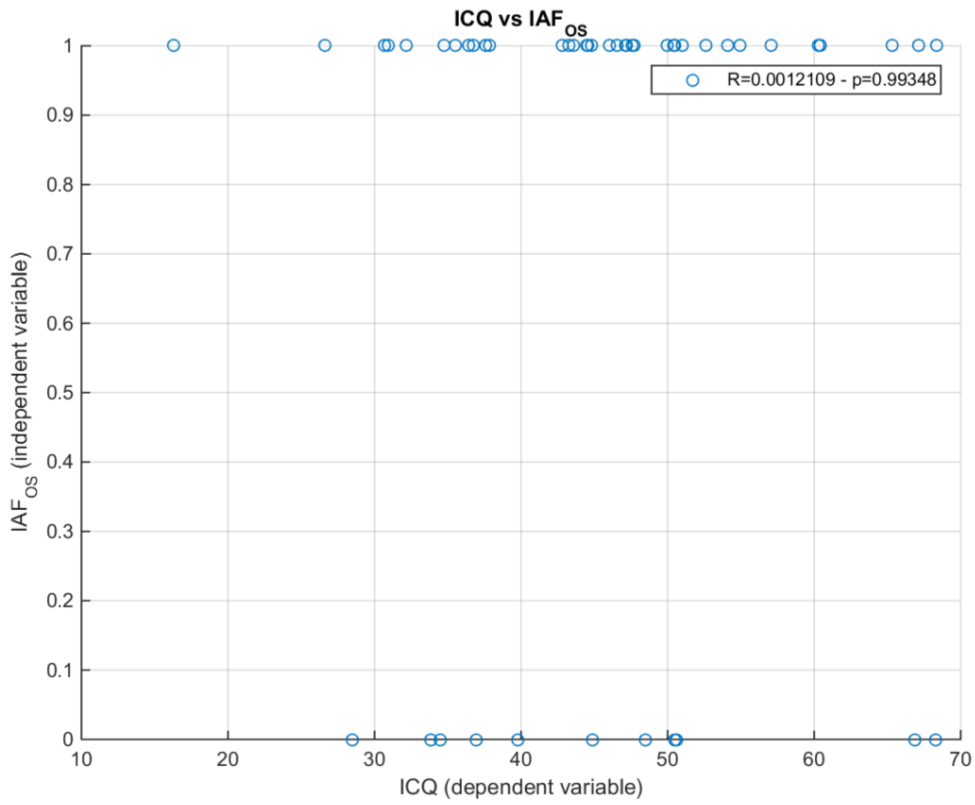

Figure 6. ICQ (dependent) Vs IAF_OS (independent) variable.

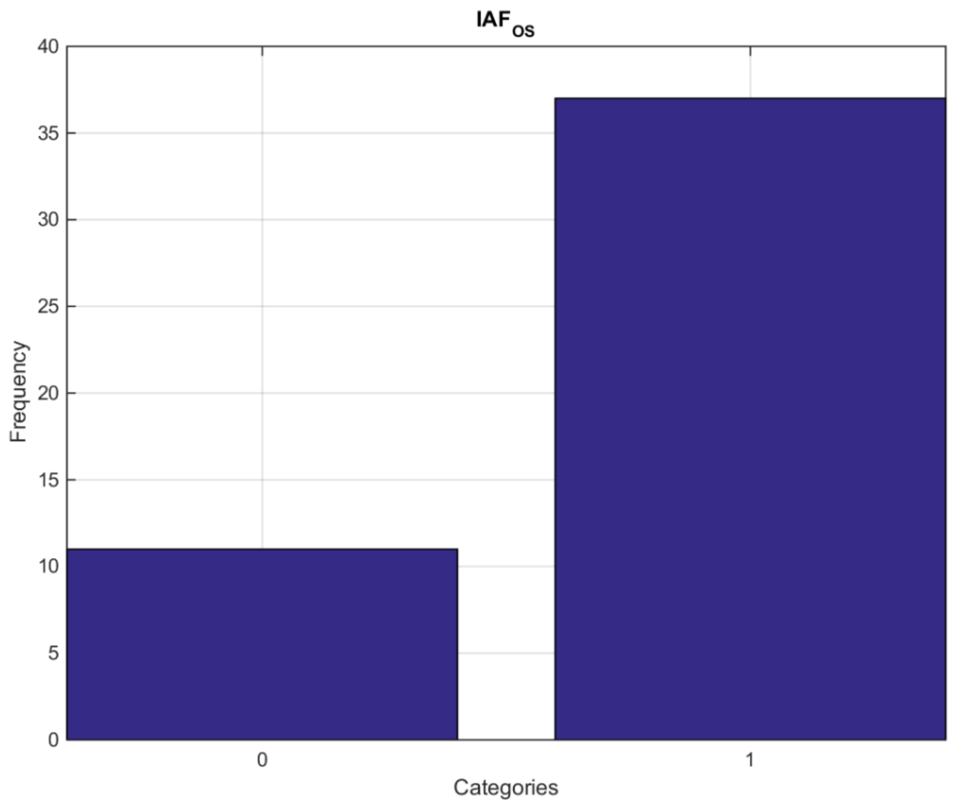

Figure 7. IAF considers functional reporting to internal auditing committee (IAF_OS).

Table 2. Descriptive statistics for IAF_OS independent variable.

IAF_OS	0	1
Frequency	11	37
Percentage (%)	22.91667	77.08333
Cumulative (%)	22.91667	100
Mean	0.770833	
Standard Deviation	0.424744	
Median	1	
Mode	1	
Variance	0.180408	
R correlation	1	0.001211
P matrix	1	0.993483
	0.993483	1

The assessment of the work relationship (WKREL), denoting whether the auditing committee reviews internal IAF program executed by internal auditors, is shown in Figure 9. Over 80% of the auditors contacted replied mostly neutrally on this matter. Moreover, as shown in Table 3, the regression coefficient (R correlation) of this variable takes a positive value, denoting that **H2** hypothesis (See section 4.2.2) is **true** with a confidence level smaller than 95% since the p-value = 0.58>0.05 corresponds to a non-significant correlation in R = 0.08 and a high probability of observing the respective null hypothesis (see Figure 8).

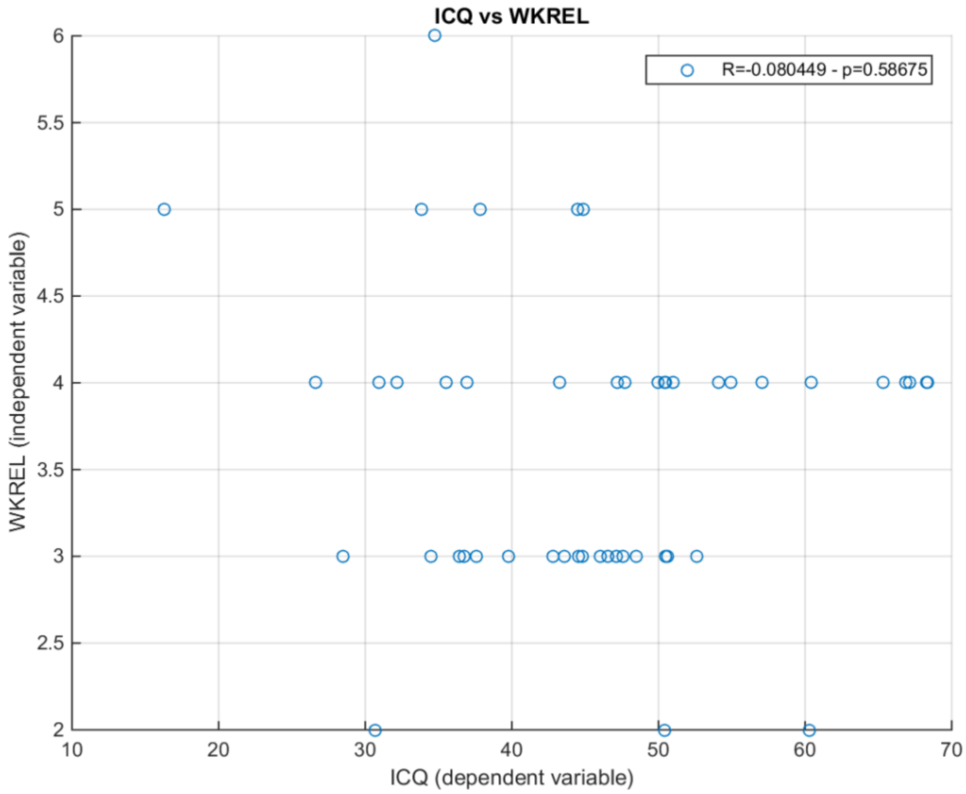

Figure 8. ICQ (dependent) Vs WKREL (independent) variable.

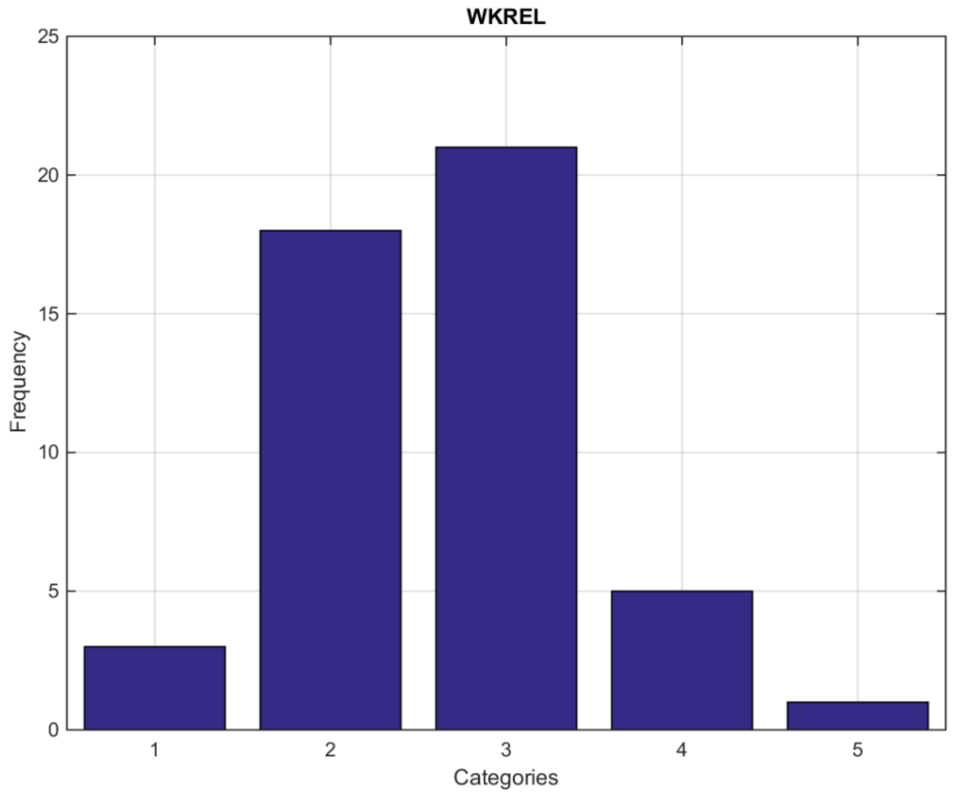

Figure 9. Internal auditing committee reviews the executed program by auditors (WKREL).

Table 3. Descriptive statistics for WKREL independent variable.

WKREL	1	2	3	4	5
Frequency	3	18	21	5	1
Percentage (%)	6.25	37.5	43.75	10.41667	2.083333
Cumulative (%)	6.25	43.75	87.5	97.91667	100
Mean	2.645833				
Standard Deviation	0.837666				
Median	4				
Mode	4				
Variance	0.701684				
R correlation	1	0.08045			
P matrix	1	0.586746			
	0.586746	1			

The assessment of the internal audit staff competence and exploitable skills, denoted as STF_COMP, is shown in Figure 11. Moreover, as shown in Table 4, the regression coefficient (R correlation) of this variable takes a negative value, denoting that **H3** hypothesis (See section 4.2.3) is **not true** with a confidence level greater than 95% since the p-value = 0.0187<0.05 corresponds to a significant correlation in R = -0.33821<0 denoting a negative association (see Figure 10).

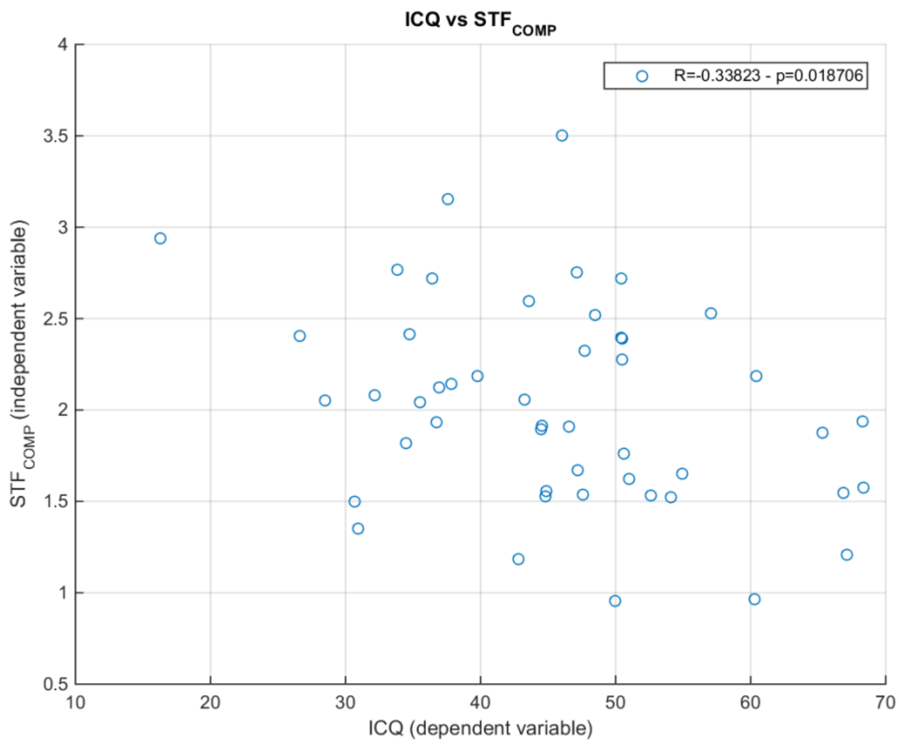

Figure 10. ICQ (dependent) Vs STF_COMP (independent) variable.

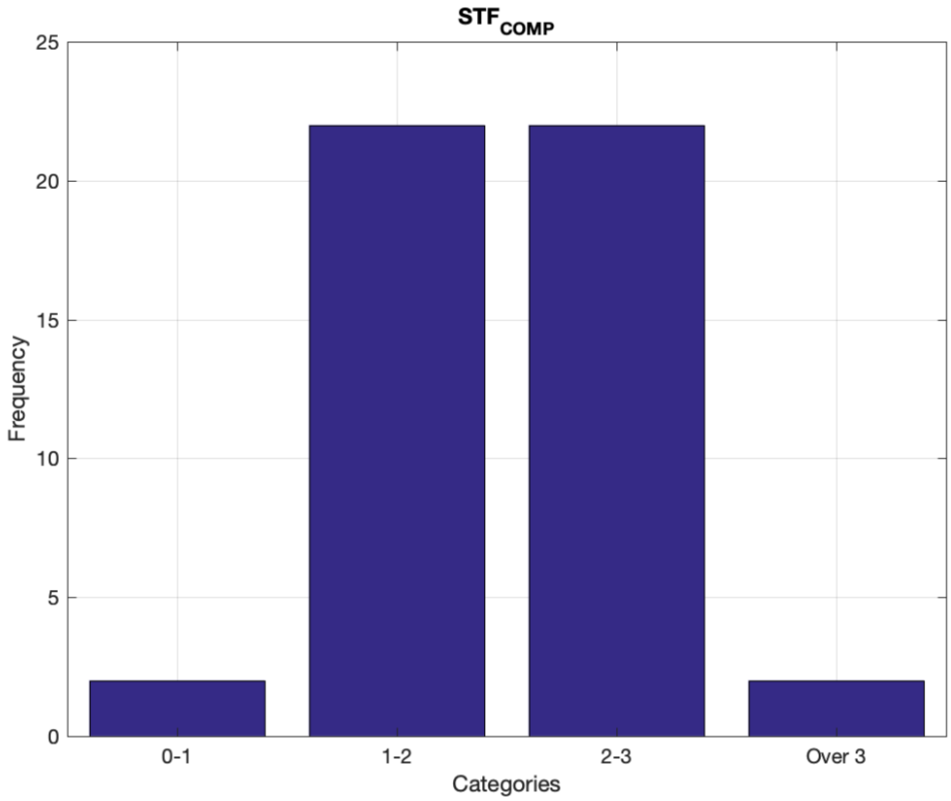

Figure 11. Assessment of the internal audit staff competence and exploitable skills (STF_COMP).

Table 4. Descriptive statistics for STF_COMP independent variable.

STF_COMP	0-1	1-2	2-3	Over 3
Frequency	2	22	22	2
Percentage (%)	4.1667	45.833	45.833	4.1667
Cumulative (%)	4.1667	50	95.833	100
Mean	2.0254			
Standard Deviation	0.5569			
Median	1.9903			
Mode	0.9565			
Variance	0.3102			
R correlation	1	-0.3382		
P matrix	1	0.0187		
	0.0187	1		

The physical variables that participate in the calculation of STF_COMP are standardized and fused as shown in section 5.3.2 above. Starting with the average years of auditors' higher education, denoted with EDU, as shown in Figure 12 and Table 5, it presents a mean value of 6.6 years and a deviation of 3 suggest a quite mediocre-to-high educational background of auditors.

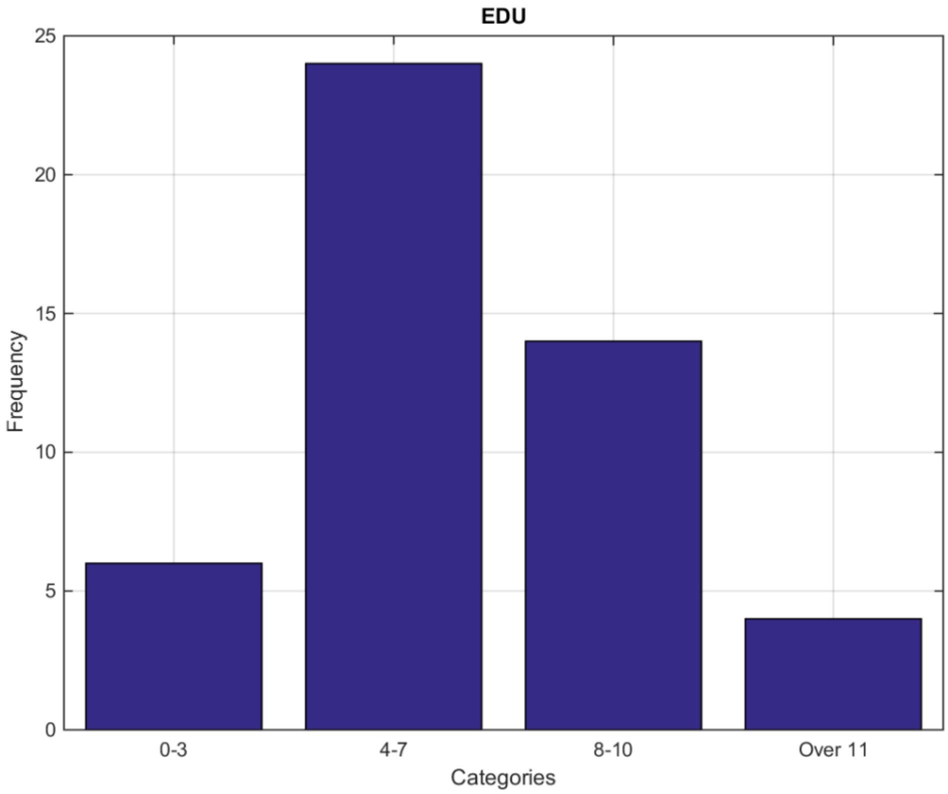

Figure 12. Average number of the auditors' higher education in years independent variable (EDU).

Table 5. Descriptive statistics for EDU independent variable.

EDU	0-3	4-7	8-10	Over 11
Frequency	6	24	14	4
Percentage (%)	12.5	50	29.16667	8.333333
Cumulative (%)	12.5	62.5	91.66667	100
Mean	6.583333			
Standard Deviation	2.95234			
Median	6			
Mode	4			
Variance	8.716312			
R correlation	1	0.058542		
P matrix	1	0.692667		
	0.692667	1		

The average years of auditors' experience, denoted with EXP, as shown in Figure 13 and Table 6, presenting a mean value of 6.6 years and a deviation of 2.9 suggest moderately, yet not amateur level, experienced groups of auditors for each company.

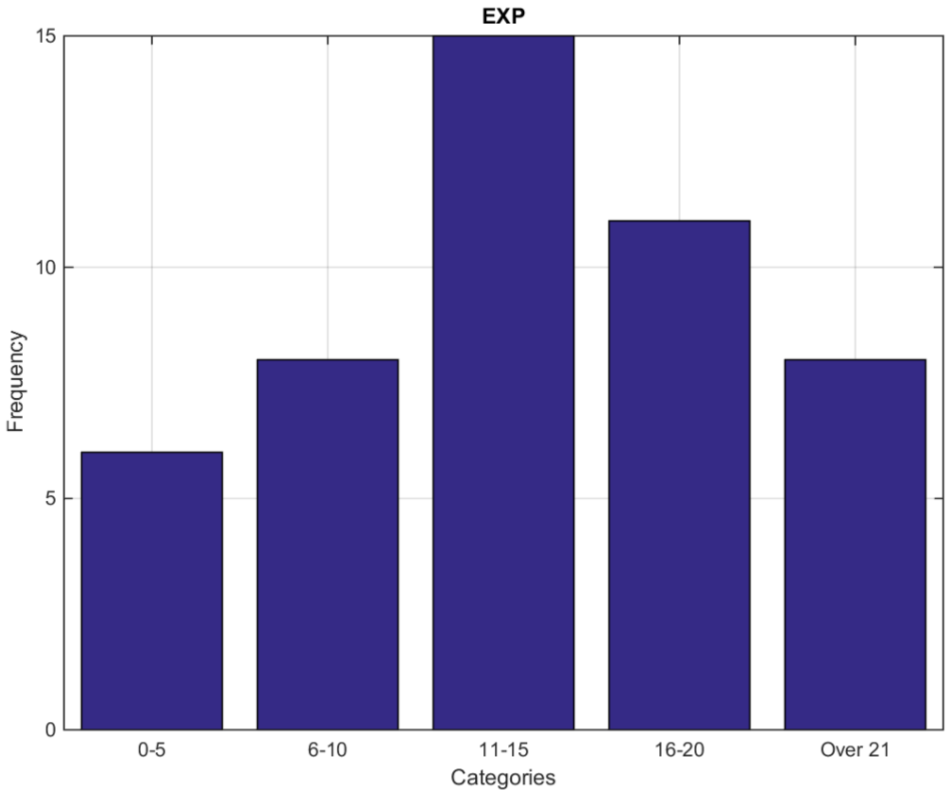

Figure 13. Average number of the auditors' experience in years independent variable (EXP).

Table 6. Descriptive statistics for EXP independent variable.

EXP	0-5	6-10	11-15	16-20	Over 21
Frequency	6	8	15	11	8
Percentage (%)	12.5	16.66667	31.25	22.91667	16.66667
Cumulative (%)	12.5	29.16667	60.41667	83.33333	100
Mean	14.0625				
Standard Deviation	6.615283				
Median	13.5				
Mode	13				
Variance	43.76197				
R correlation	1	-0.33058			
P matrix	1	0.02174			
	0.02174	1			

The certification index, denoted with CERT, which is a fraction of the total number of certified auditors (mean=3.25, std=1.73), as shown in Figure 14 and Table 7, over the number of total ones (mean=10.7, std=3.4), as shown in Figure 15 and Table 8.

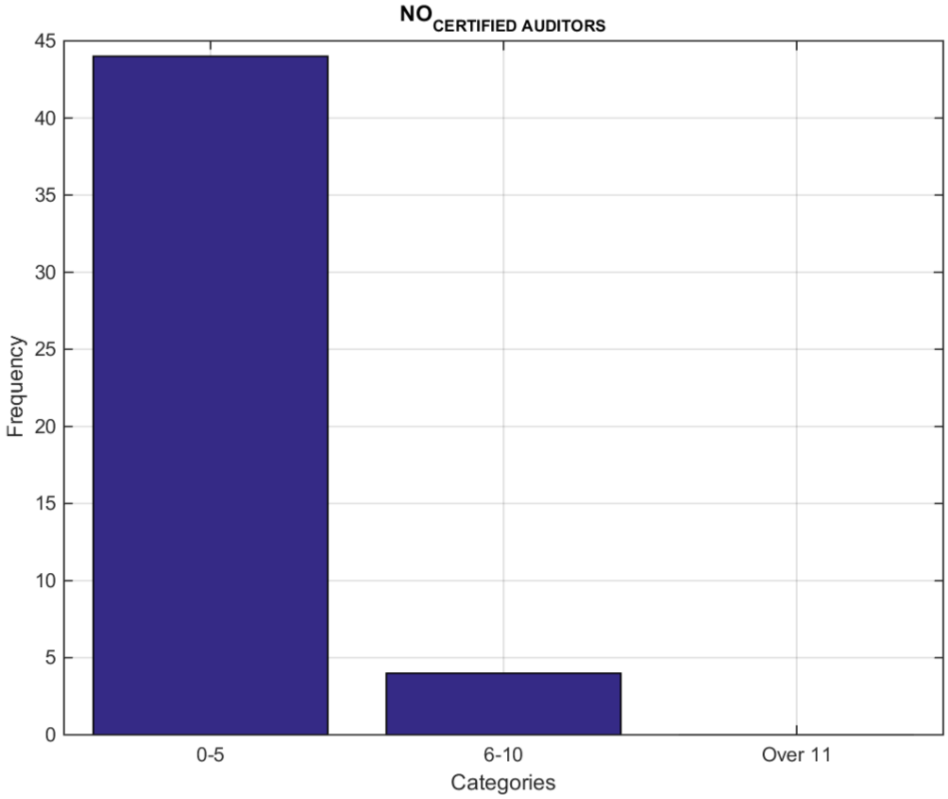

Figure 14. Number of auditors with a formal certification (NO_CERTIFIED_AUDITORS).

Table 7. Descriptive statistics for NO_CERTIFIED_AUDITORS independent variable.

NO_CERTIFIED AUDITORS	0-5	6-10	Over 11
Frequency	44	4	0
Percentage (%)	91.66667	8.333333	0
Cumulative (%)	91.66667	100	100
Mean	3.25		
Standard Deviation	1.732051		
Median	3		
Mode	3		
Variance	3		
R correlation	1	0.059898	
P matrix	1	0.685912	
	0.685912	1	

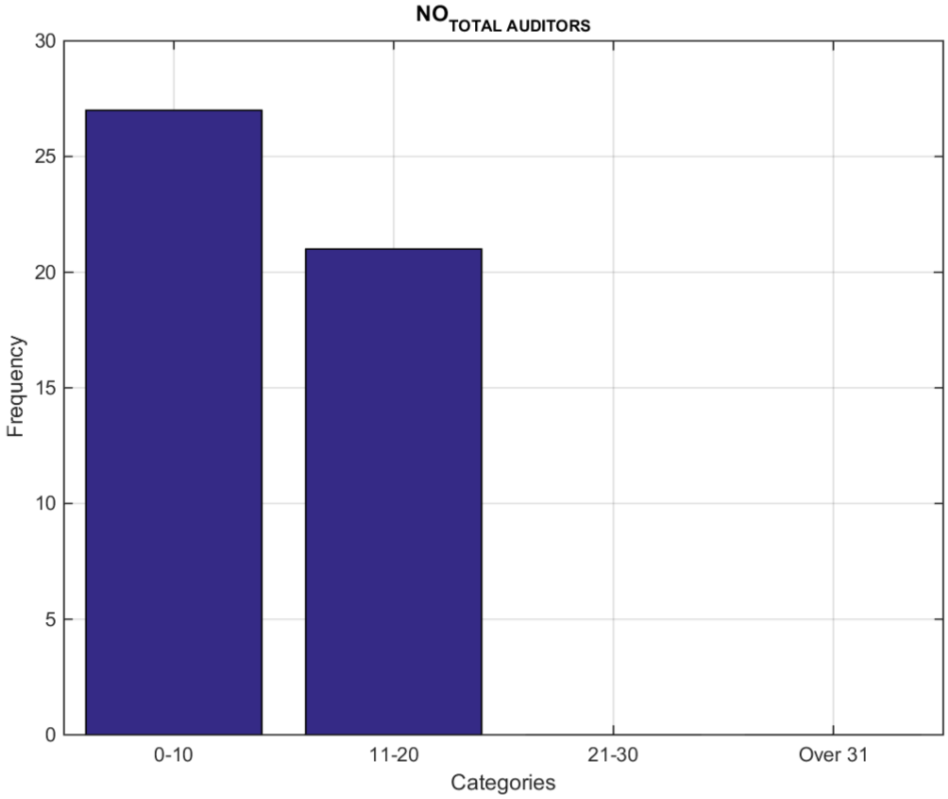

Figure 15. Total number of auditors (NO_TOTAL_AUDITORS).

Table 8. Descriptive statistics for NO_TOTAL_AUDITORS independent variable.

NO_TOTAL AUDITORS	0-10	11-20	21-30	Over 31
Frequency	27	21	0	0
Percentage (%)	56.25	43.75	0	0
Cumulative (%)	56.25	100	100	100
Mean	10.72917			
Standard Deviation	3.381675			
Median	10			
Mode	10			
Variance	11.43573			
R correlation	1	0.223524		
P matrix	1	0.126714		
	0.126714	1		

The hours of annual training, denoted with TRN, for internal auditors' skills evolution and enhancement (mean=8.9, std=4.1), as shown in Figure 16 and Table 9.

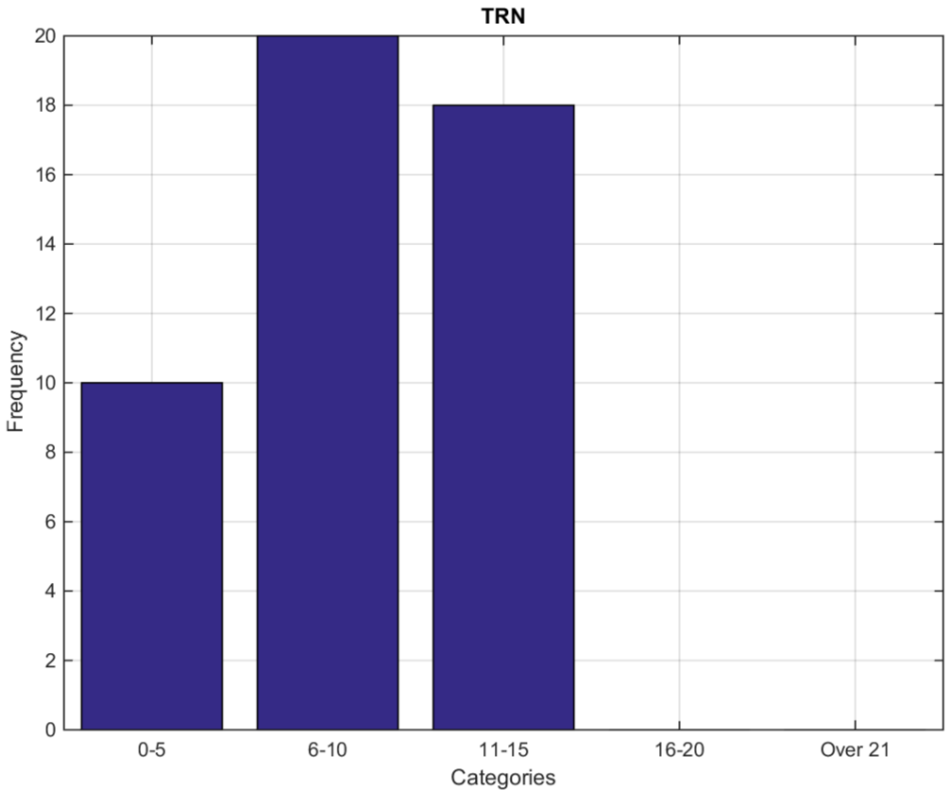

Figure 16. Number of average training hours per year spent by auditors (TRN).

Table 9. Descriptive statistics for TRN independent variable.

TRN	0-5	6-10	11-15	16-20	Over 21
Frequency	10	20	18	0	0
Percentage (%)	20.83333	41.66667	37.5	0	0
Cumulative (%)	20.83333	62.5	100	100	100
Mean	8.875				
Standard Deviation	4.108243				
Median	9.5				
Mode	10				
Variance	16.87766				
R correlation	1	-0.37051			
P matrix	1	0.009535			
	0.009535	1			

The investment in IAF execution, which refers to the available human resources occupied for such a purpose within the company, denoted with IAF_HMR (mean=5, std=2.2), as shown in Figure 19 and Table 11. The collected values are then parsed through the natural logarithm function to form the IAF_INV independent variable as shown in Figure 18. Moreover, as shown in Table 10, the regression coefficient (R correlation) of this variable takes a negative value, denoting that **H4** hypothesis (See section 4.2.4) is **true** with a confidence level smaller than 95% since the p-value = 0.7268>0.05 corresponds to a non-significant correlation in R = 0.0518 and a high probability of observing the respective null hypothesis (Figure 17).

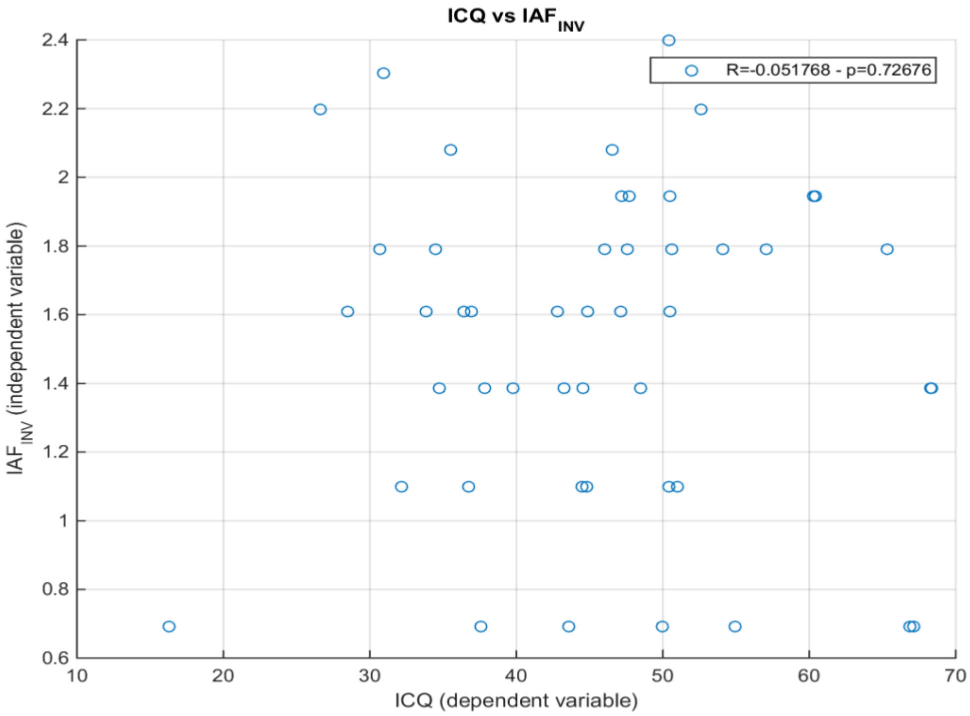

Figure 17. ICQ (dependent) Vs IAF_INV (independent) variable

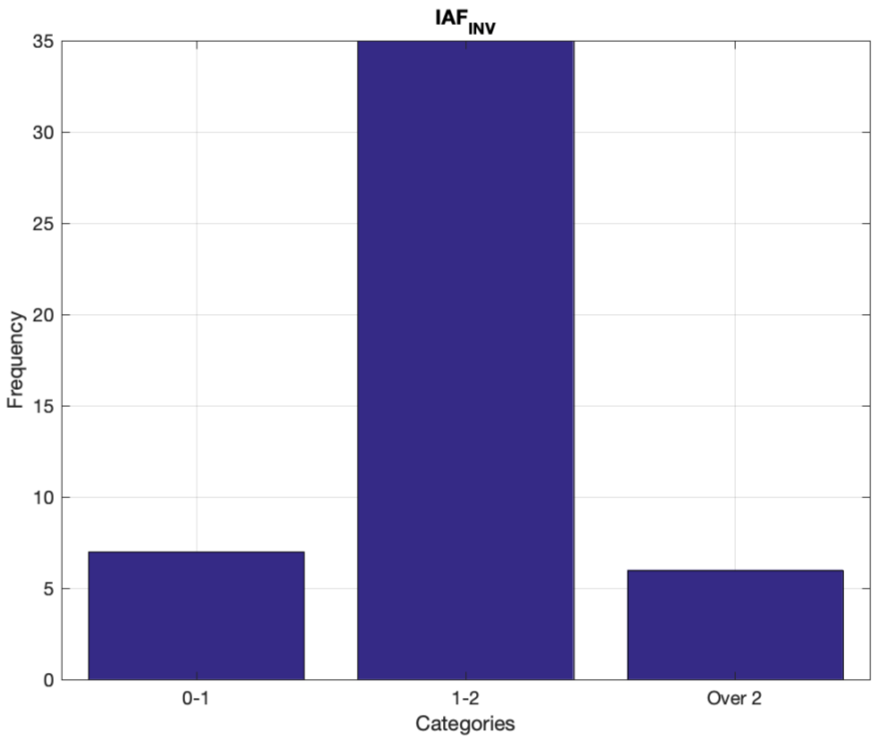

Figure 18. Natural logarithm of IAF_HMR independent variable (IAF_INV).

Table 10. Descriptive statistics for IAF_INV independent variable.

IAF_INV	0-1	1-2	Over 2
Frequency	7	35	6
Percentage (%)	14.5833	72.9167	12.5
Cumulative (%)	14.5833	87.5	100
Mean	1.5151		
Standard Deviation	0.4722		
Median	1.6094		
Mode	1.3863		
Variance	0.2230		
R correlation	1	0.0518	
P matrix	1	0.7268	
	0.7268	1	

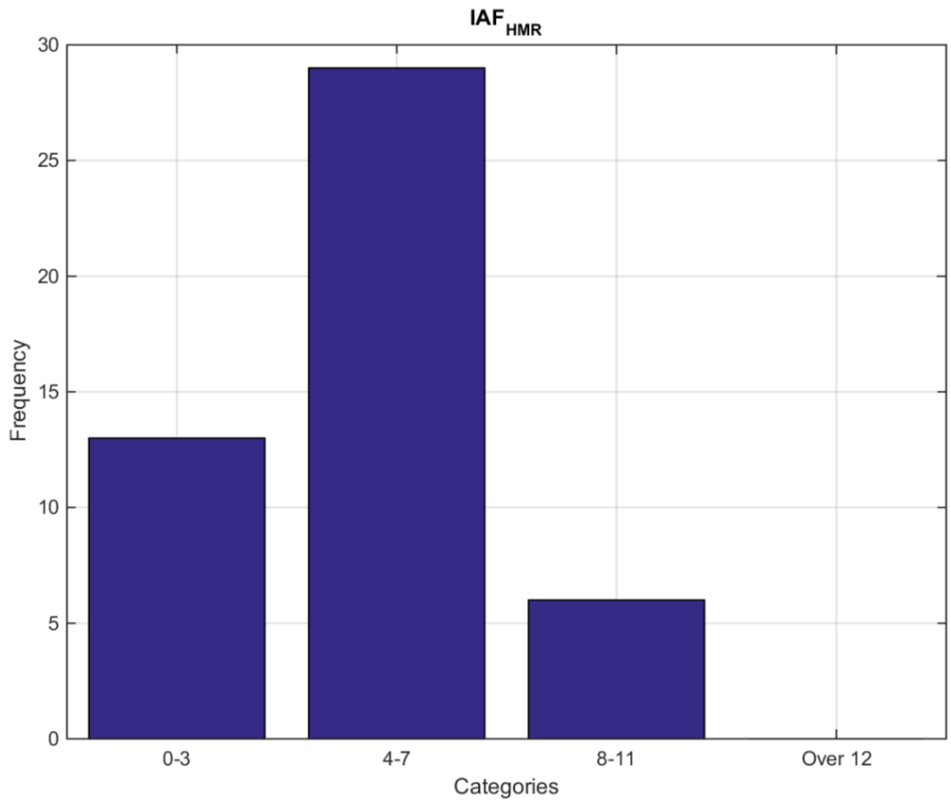

Figure 19. Number of people involved within the execution of the IAF program (IAF_HMR).

Table 11. Descriptive statistics for IAF_HMR independent variable.

IAF_HMR	0-3	4-7	8-11	Over 12
Frequency	13	29	6	0
Percentage (%)	27.08333	60.41667	12.5	0
Cumulative (%)	27.08333	87.5	100	100

Mean	5.041667			
Standard Deviation	2.230908			
Median	5			
Mode	4			
Variance	4.97695			
R correlation	1	0.05456		
P matrix	1	0.712632		
	0.712632	1		

Regarding the quality assurance program, denoted as QAS, it is calculated using the dichotomized (by the respective median value). Moreover, as shown in Figure 21 and in Table 12 the regression coefficient (R correlation) of this variable takes a negative value, denoting that **H5** hypothesis (See section 4.2.5) is **true** with a confidence level smaller than 95% since the p-value = 0.8318>0.05 corresponds to a significant correlation in R = 0.0315 and a low probability of observing the respective null hypothesis (Figure 20).

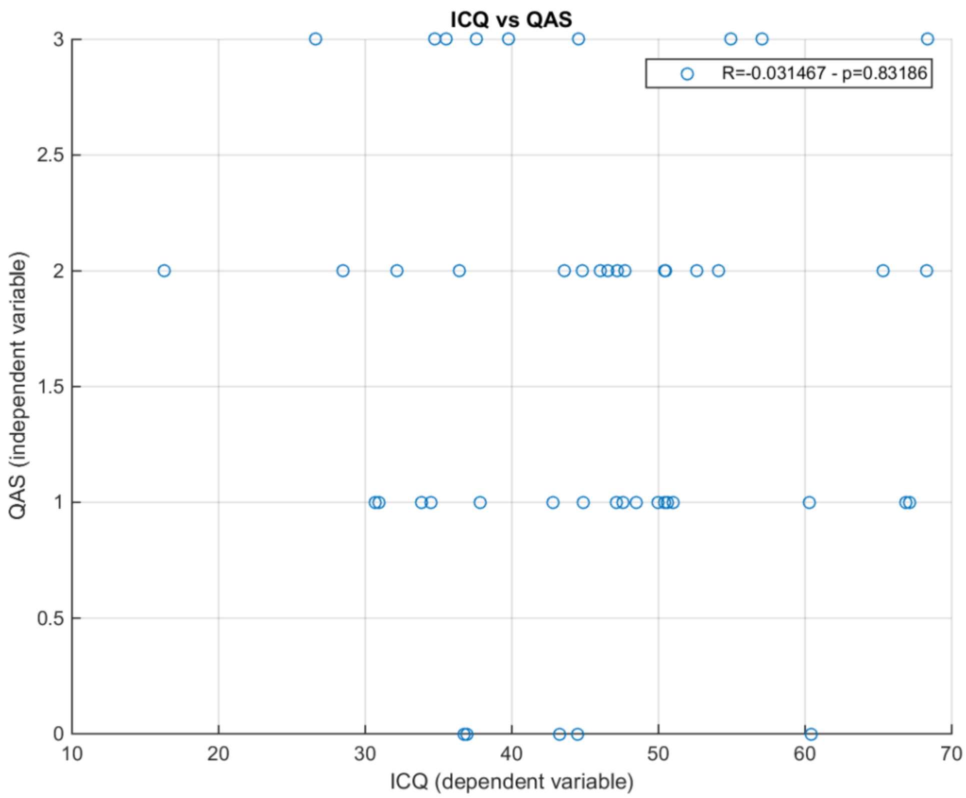

Figure 20. ICQ (dependent) Vs QAS (independent) variable

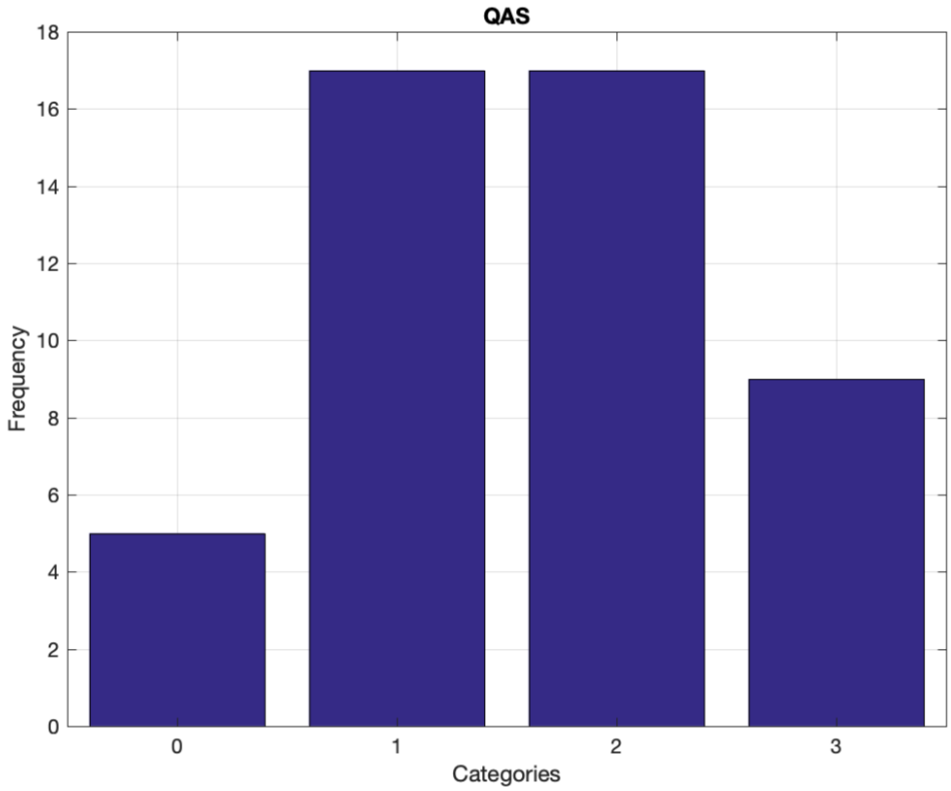

Figure 21. Aggregated quality assurance program.

Table 12. Descriptive statistics for QAS independent variable.

QAS	0	1	2	3
Frequency	5	17	17	9
Percentage (%)	10.4167	35.4167	35.4167	18.75
Cumulative (%)	10.4167	45.833	81.25	100
Mean	1.625			
Standard Deviation	0.9138			
Median	2			
Mode	1			
Variance	0.8351			
R correlation	1	0.0315		
P matrix	1	0.8318		
	0.8318	1		

The following physical variables are collected by the usable sampled questionnaires to calculate the QAS:

a) The existence of a QA program, denoted with QAPX, which reflects whether a QA program exists or not within the sampled firm's operational procedures (mean=0.5, std=0.5), as shown in Figure 22 and Table 13.

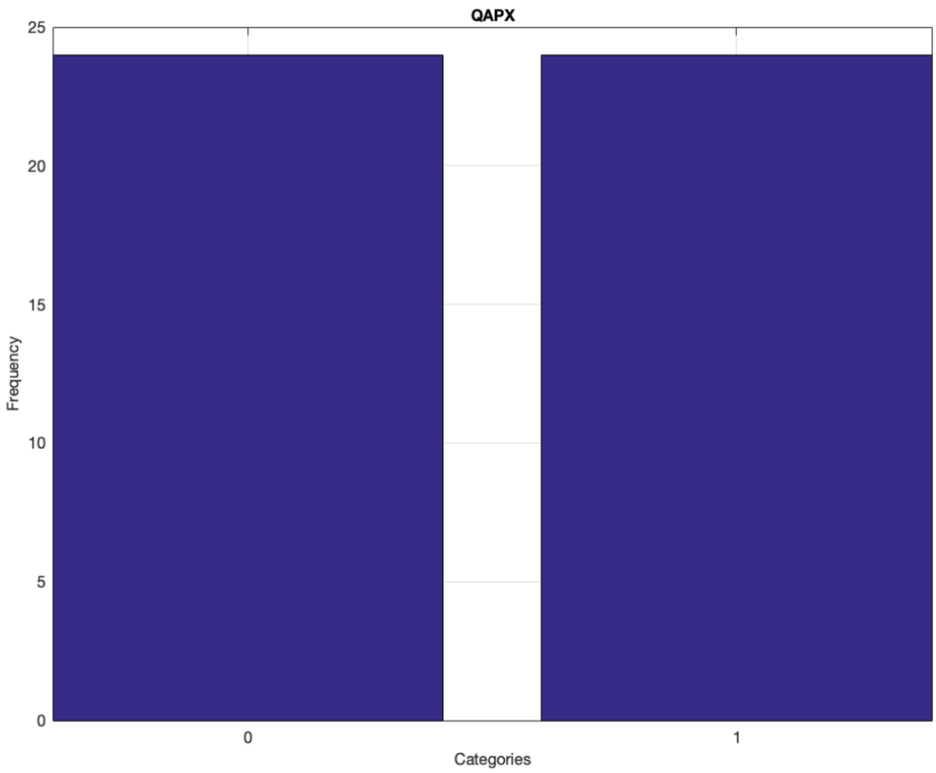

Figure 22. The existence (or not) of a Quality Assurance program within the operational procedures of the firm (QAPX).

Table 13. Descriptive statistics for QAPX independent variable.

QAPX	0	1
Frequency	24	24
Percentage (%)	50	50
Cumulative (%)	50	100
Mean	0.5	
Standard Deviation	0.505291	
Median	0.5	
Mode	0	
Variance	0.255319	
R correlation	1	0.005901
P matrix	1	0.968247
	0.968247	1

b) The internal assessment index, denoted with NT_ASS, which is the normalized average of the level of utilization of internal continuous monitoring tools (UT_MON, mean=2.97, std=1.13) as shown in Figure 23 and Table 14, and the reporting tendency of periodic auditing reviews (R_TEND, mean=3.02, std=1.15), as shown in Figure 24 and Table 15.

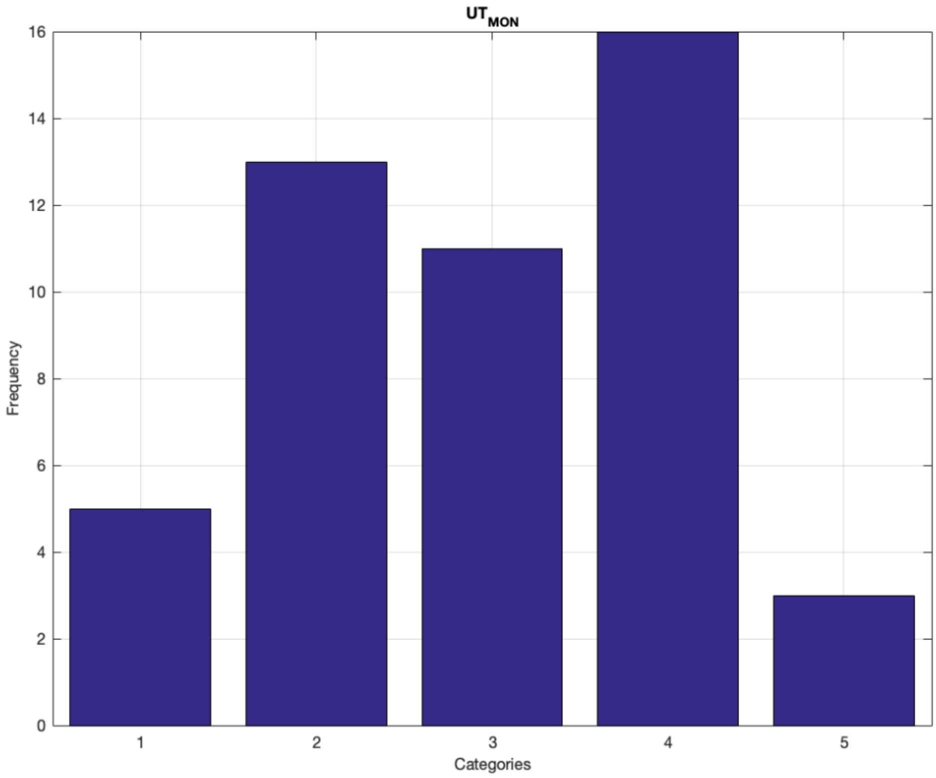

Figure 23. The level of internal continuous monitoring tools utilization (UT_MON).

Table 14. Descriptive statistics for UT_MON independent variable.

UT_MON	1	2	3	4	5
Frequency	5	13	11	16	3
Percentage (%)	10.41667	27.08333	22.91667	33.33333	6.25
Cumulative (%)	10.41667	37.5	60.41667	93.75	100
Mean	2.979167				
Standard Deviation	1.139047				
Median	3				
Mode	4				
Variance	1.297429				
R correlation	1	0.06409			
P matrix	1	0.665182			
	0.665182	1			

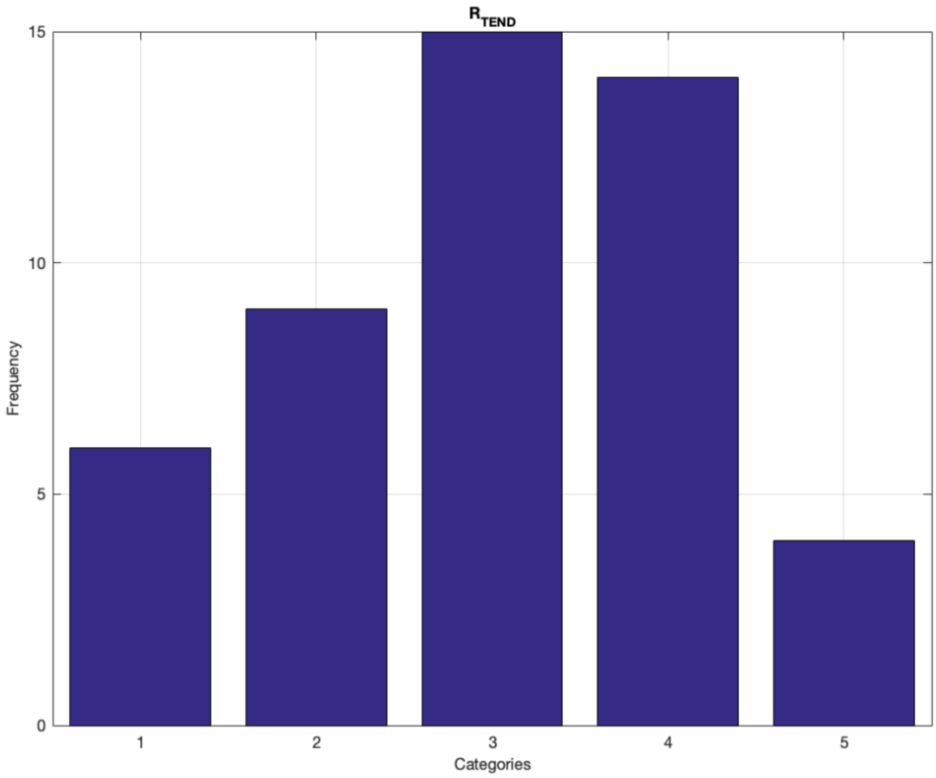

Figure 24. The level of periodic internal auditing reporting tendency (R_TEND).

Table 15. Descriptive statistics for R_TEND independent variable.

R_TEND	1	2	3	4	5
Frequency	6	9	15	14	4
Percentage (%)	12.5	18.75	31.25	29.16667	8.333333
Cumulative (%)	12.5	31.25	62.5	91.66667	100
Mean	3.020833				
Standard Deviation	1.157576				
Median	3				
Mode	3				
Variance	1.339982				
R correlation	1	0.071415			
P matrix	1	0.629554			
	0.629554	1			

c) The external assessment, denoted with XT_ASS, calculated as the average of: a) the existence of an external quality assessment (EQXA, mean=0.5, std=0.5) as shown in Figure 25 and Table 16; b) the implementation of a fully external assessment or self-assessment assisted by external validation (EXASS, mean=0.625, std=0.489) as shown in Figure 26 and Table 17, and; c) the periodic implementation of an internal auditing external evaluation every five years (PREX, mean=0.416, std=0.498) as shown in Figure 27 and Table 18.

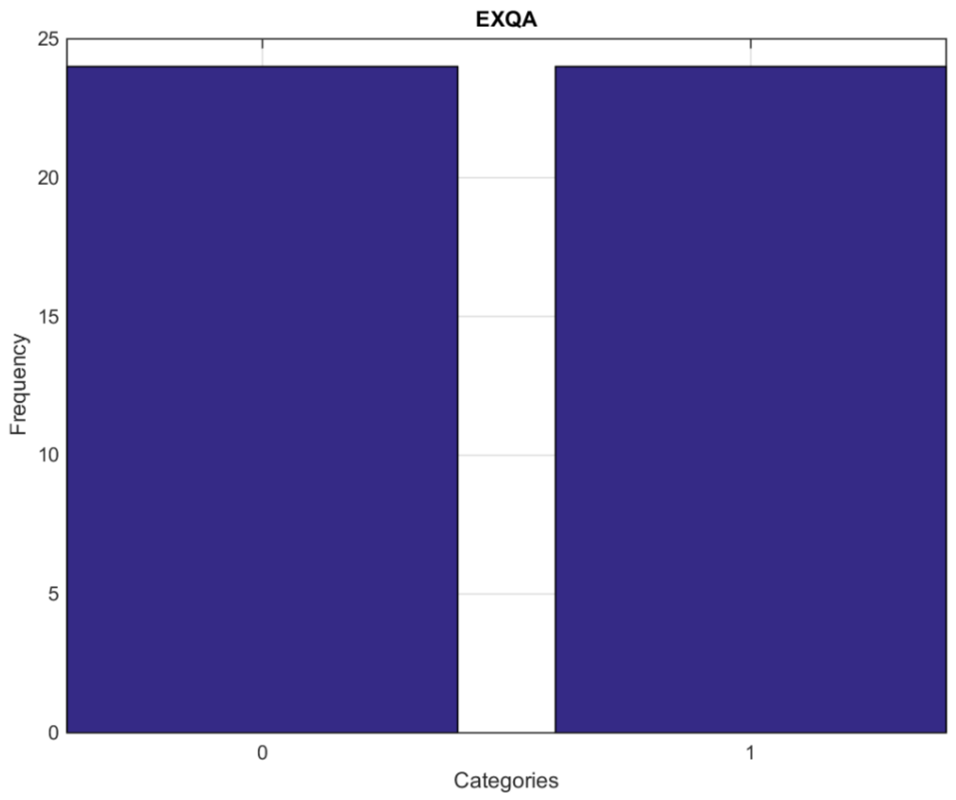

Figure 25. The existence of external quality assessment programs (EXQA).

Table 16. Descriptive statistics for EXQA independent variable.

EXQA	0	1
Frequency	24	24
Percentage (%)	50	50
Cumulative (%)	50	100
Mean	0.5	
Standard Deviation	0.505291	
Median	0.5	
Mode	0	
Variance	0.255319	
R correlation	1	-0.14819
P matrix	1	0.314806
	0.314806	1

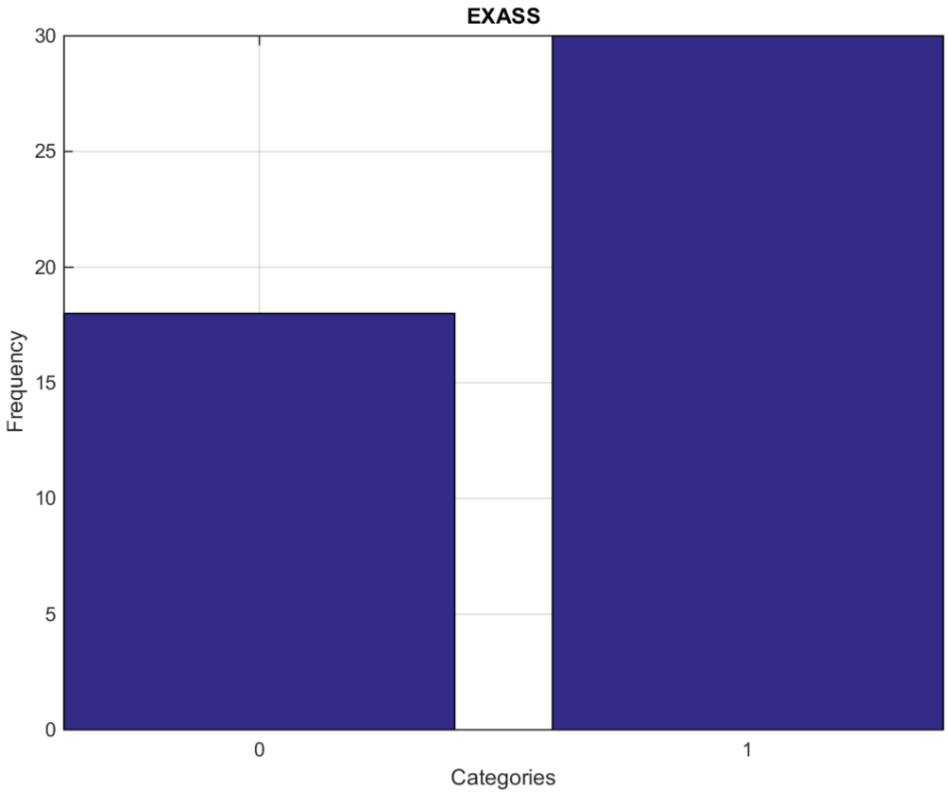

Figure 26. The implementation of a fully external assessment (1) or self-assessment assisted by external validation (2) (EXASS).

Table 17. Descriptive statistics for EXASS independent variable.

EXASS	0	1
Frequency	18	30
Percentage (%)	37.5	62.5
Cumulative (%)	37.5	100
Mean	0.625	
Standard Deviation	0.489246	
Median	1	
Mode	1	
Variance	0.239362	
R correlation	1	-0.03567
P matrix	1	0.809809
	0.809809	1

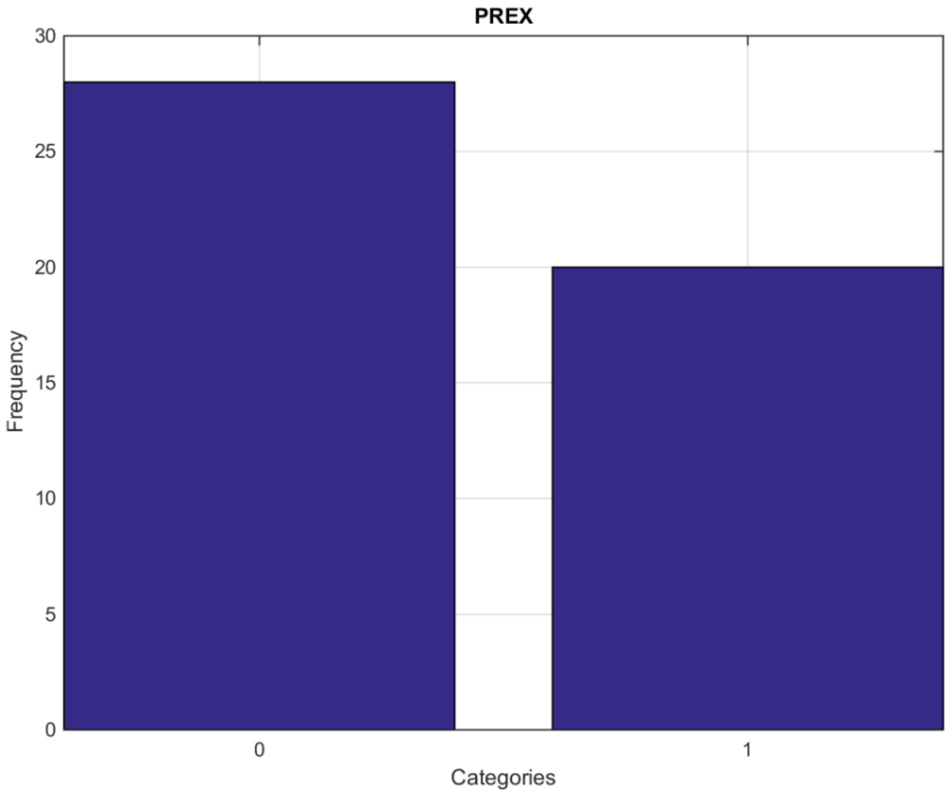

Figure 27. The periodic external evaluation of internal auditing every five years (PREX).

Table 18. Descriptive statistics for PREX independent variable.

PREX	0	1
Frequency	28	20
Percentage (%)	58.33333	41.66667
Cumulative (%)	58.33333	100
Mean	0.416667	
Standard Deviation	0.498224	
Median	0	
Mode	0	
Variance	0.248227	
R correlation	1	0.180289
P matrix	1	0.220111
	0.220111	1

Finally, the follow-up on internal control deficiencies, denoted as FUP_DEF, is a post-designed dummy variable denoting whether IAF builds upon the knowledge acquired from previously observed internal control deficiencies or not (mean=0.35, std=0.48), as shown in Figure 29. Moreover, as shown in Table 19, the regression coefficient (R correlation) of this variable takes a positive value, denoting that **H6** hypothesis (see section 4.2.6) is **true** with a confidence level smaller than 95% since the p-value =

0.658>0 corresponds to a non-significant correlation in R = 0.065 and a high probability of observing the respective null hypothesis (see Figure 28).

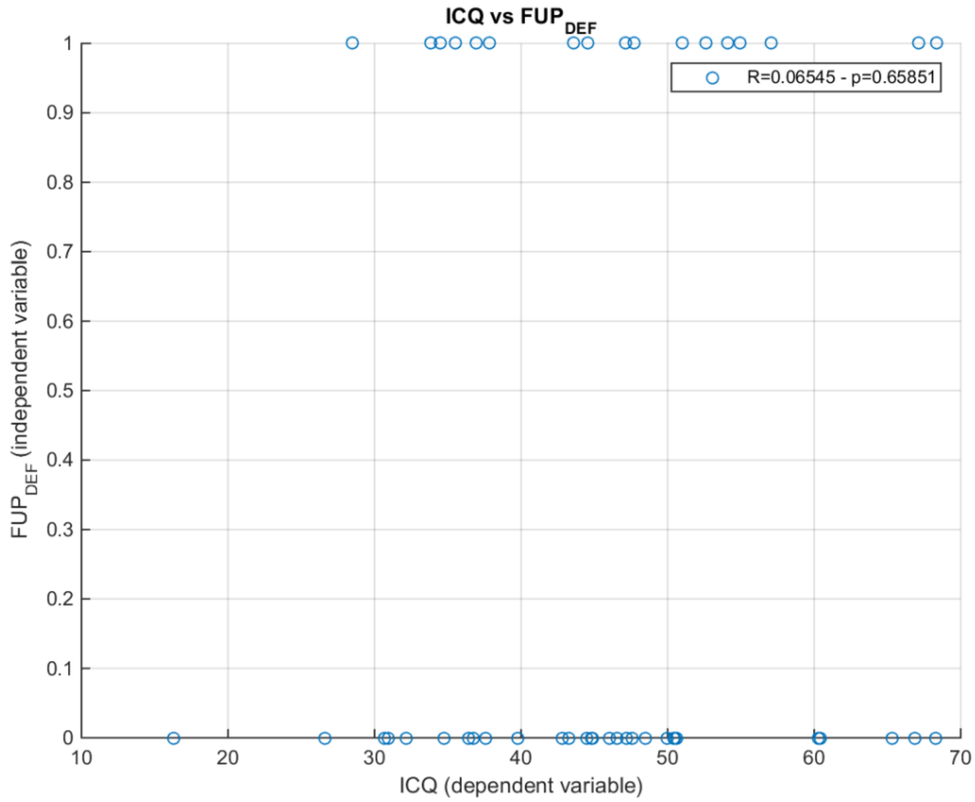

Figure 28. ICQ (dependent) Vs FUP_DEF (independent) variable

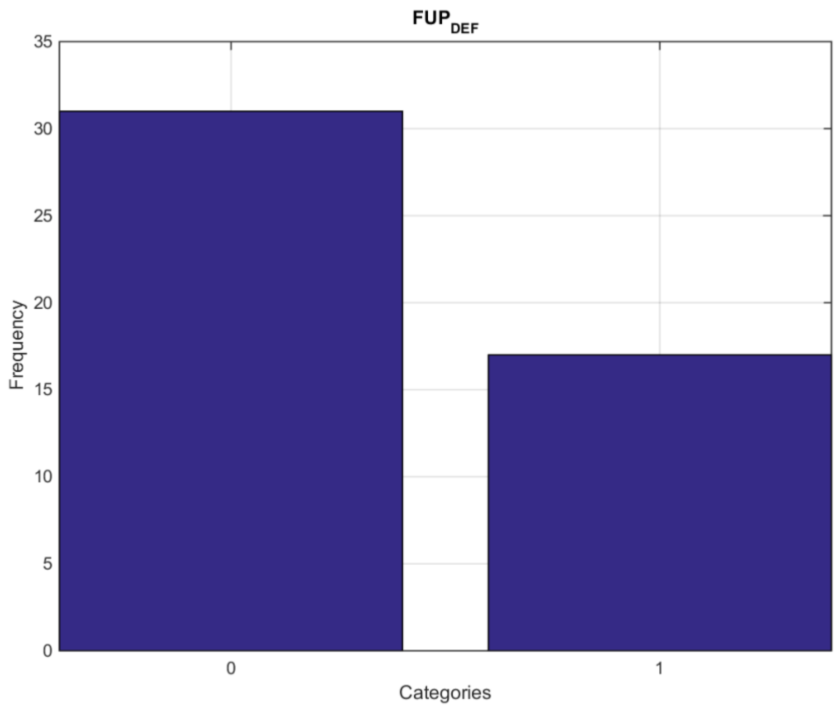

Figure 29. Whether the knowledge acquired from previously observed internal control deficiencies is being exploited or not (FUP_DEF).

Table 19. Descriptive statistics for FUP_DEF independent variable.

FUP_DEF	0	1
Frequency	31	17
Percentage (%)	64.58333	35.41667
Cumulative (%)	64.58333	100
Mean	0.354167	
Standard Deviation	0.483321	
Median	0	
Mode	0	
Variance	0.233599	
R correlation	1	0.06545
P matrix	1	0.658509
	0.658509	1

6.2.2. The control variables

The current subsection will present the descriptive statistics results of the selected control variables considered within the regression model, as presented in section 5.3.3 and 5.4. The percentage of financial experts in the audit committee (FC_XP) is a control variable indicating the background expertise and capacity of the committee members. It is derived by the fraction of the absolute number of financial experts in the audit committee "FC_XP no" results (mean=2.83, std=0.72), shown in Figure 32, over the absolute total number of audit committee members "FC_XP tot" (mean=6.03, std=1.01), shown in Figure 33. Moreover, as shown in Table 20 and Table 21, the regression coefficient (R correlation) of these variables takes a positive value, with a confidence level smaller than 95% since the p-value = 0.658>0.05 and p-value = 0.332>0.05 correspond to non-significant correlations in R = 0.065 (see Figure 30) and R = 0.143 (see Figure 31) respectively.

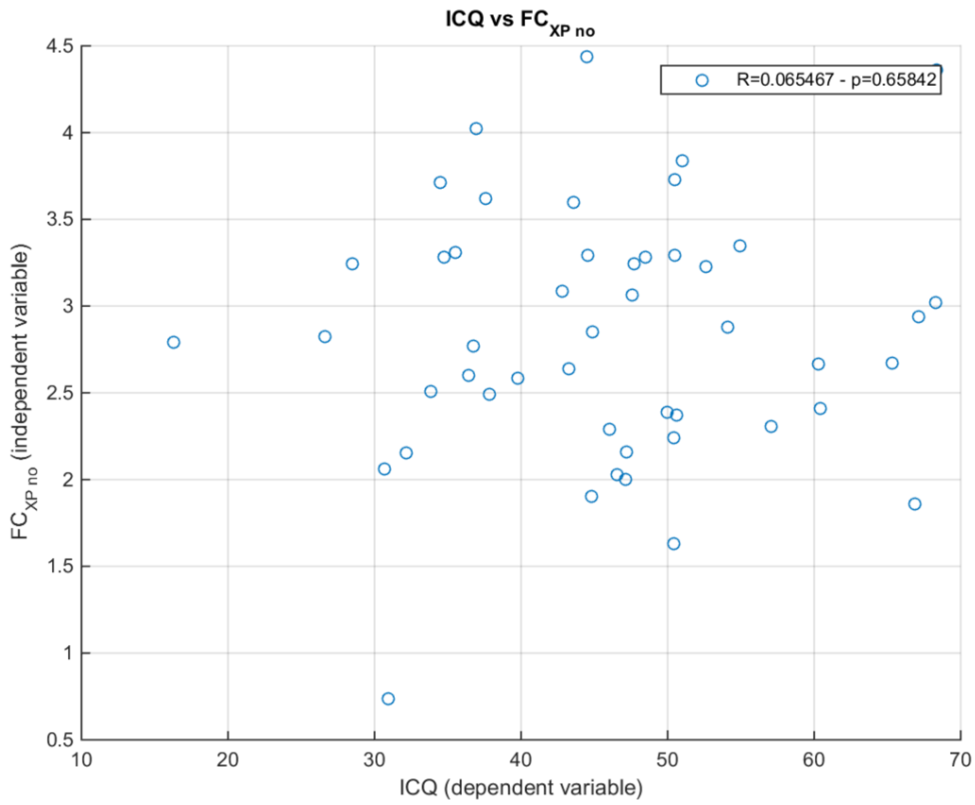

Figure 30. ICQ (dependent) Vs FC_XP_no (control) variable.

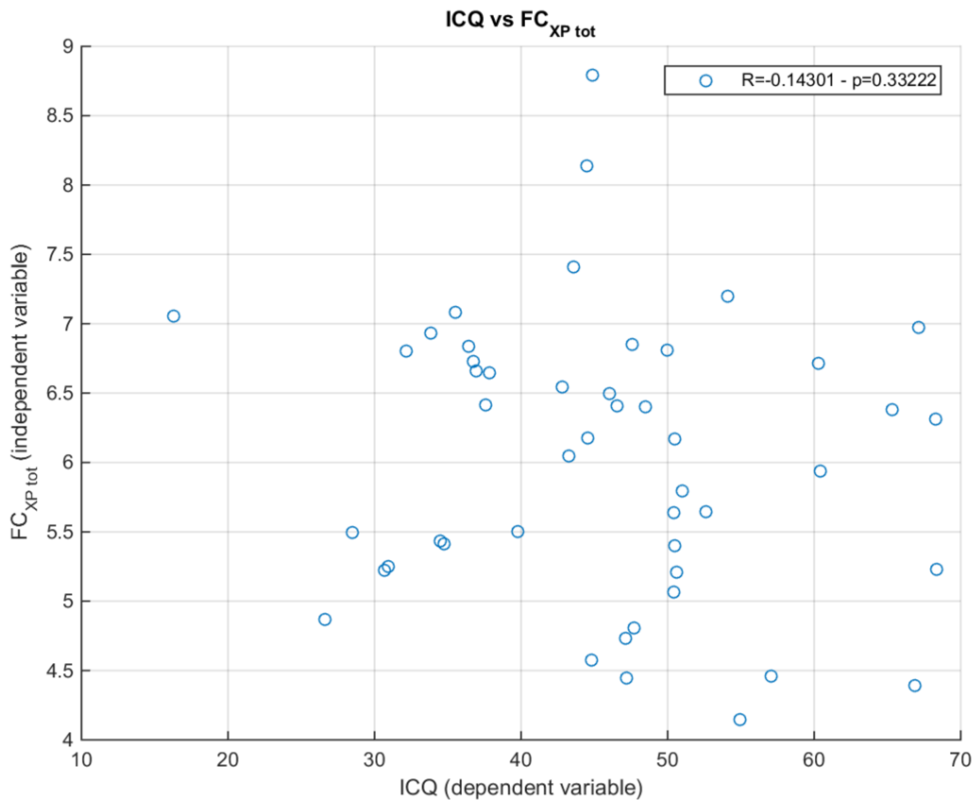

Figure 31. ICQ (dependent) Vs FC_XP_tot (control) variable

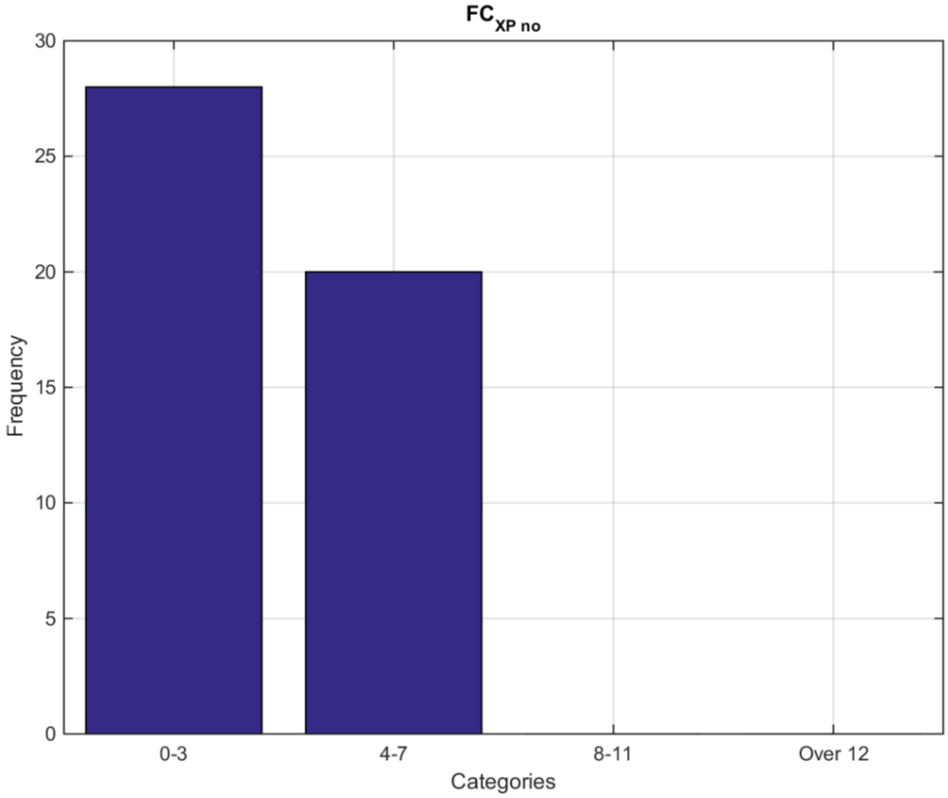

Figure 32. Number of financial experts in the audit committee (FC XP no).

Table 20. Descriptive statistics for FC_XP_no control variable.

FC_XP no	0-3	4-7	8-11	Over 12
Frequency	28	20	0	0
Percentage (%)	58.33333	41.66667	0	0
Cumulative (%)	58.33333	100	100	100
Mean	2.828428			
Standard Deviation	0.722455			
Median	2.808494			
Mode	0.737574			
Variance	0.521941			
R correlation	1	0.065467		
P matrix	1	0.658425		
	0.658425	1		

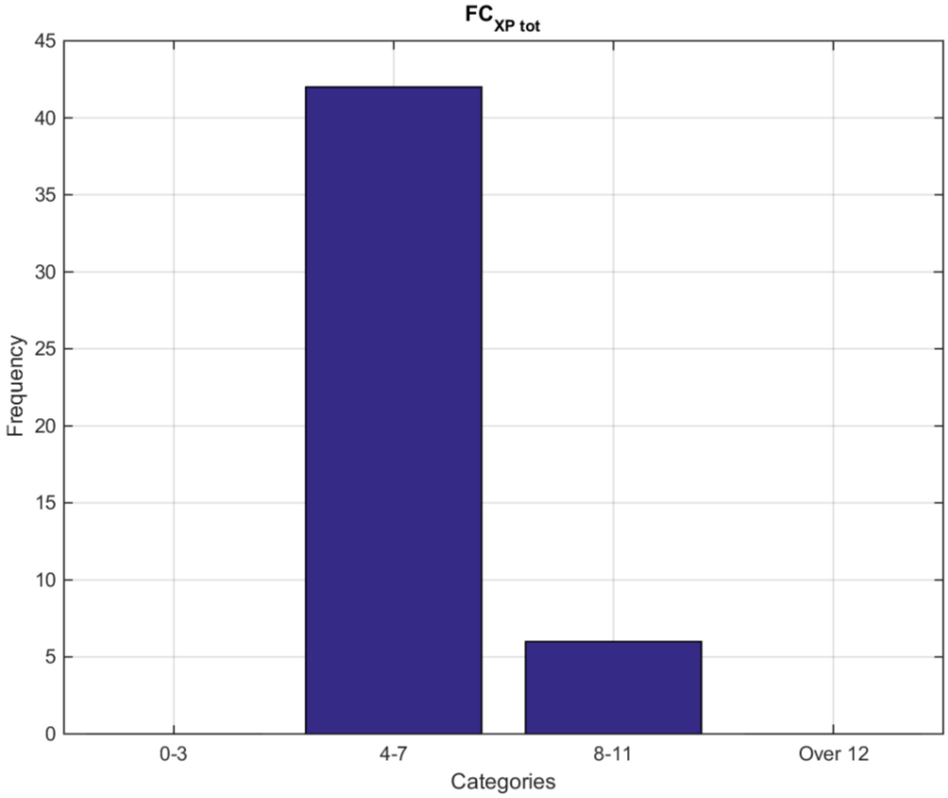

Figure 33. Number of members in the audit committee (FC XP tot).

Table 21. Descriptive statistics for FC_XP_tot control variable.

FC_XP tot	0-3	4-7	8-11	Over 12
Frequency	0	42	6	0
Percentage (%)	0	87.5	12.5	0
Cumulative (%)	0	87.5	100	100
Mean	6.034002			
Standard Deviation	1.010641			
Median	6.172417			
Mode	4.146992			
Variance	1.021396			
R correlation	1	-0.14301		
P matrix	1	0.332221		
	0.332221	1		

The natural logarithm of the entity’s sales size (LN_SLS) is a control variable indicating the size of the annual sales incoming stream of the company (SLS_SIZE). The annual sales size (SLS_SIZE) results (mean=57,268,107€; std=33,330,673€) is shown in Figure 33. Moreover, as shown in Table 22, the regression coefficient (R correlation) of these variables takes a positive value, with a confidence level smaller than 95% since the p-value = 0.513>0.05 corresponds to a non-significant correlation in R = 0.0967 (see Figure 34).

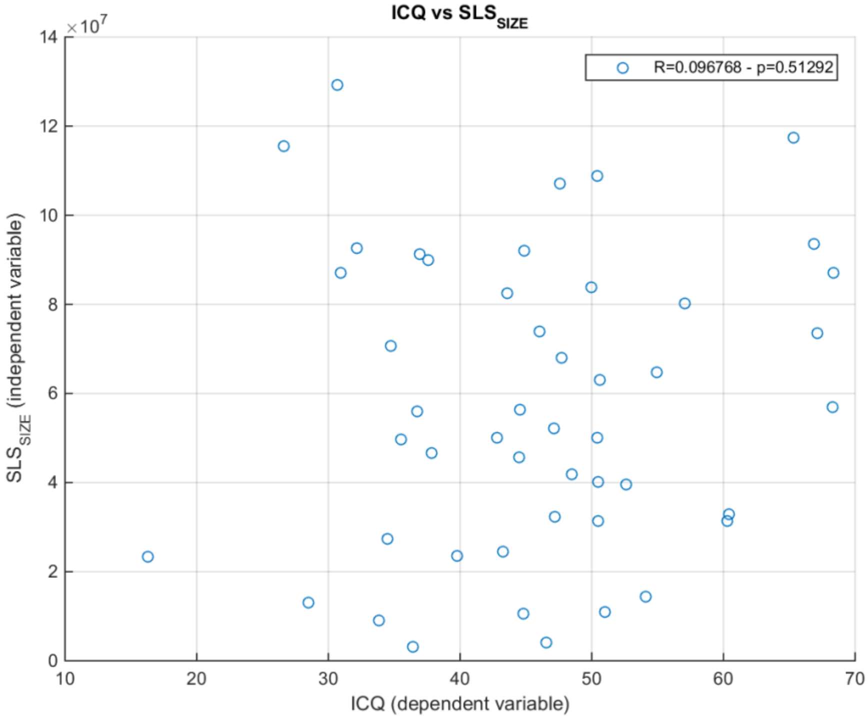

Figure 34. ICQ (dependent) Vs SLS_SIZE (control) variable.

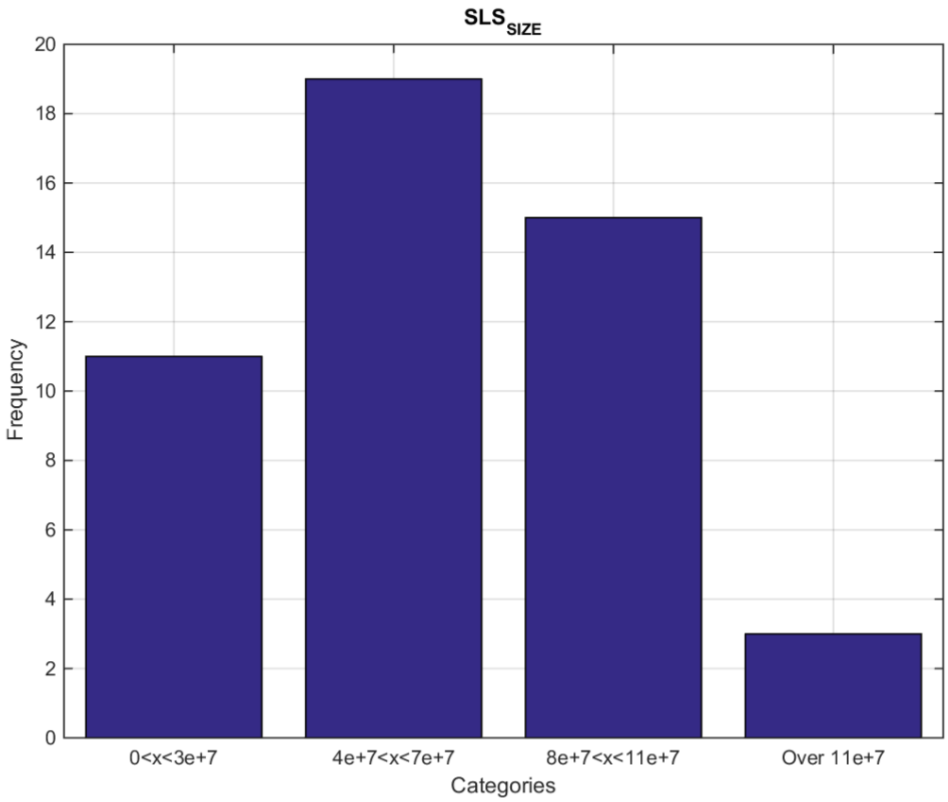

Figure 35. Size of sales incoming stream (SLS_SIZE).

Table 22. Descriptive statistics for SLS_SIZE control variable.

SLS_SIZE	0<x<3e+7	4e+7<x<7e+7	8e+7<x<11e+7	Over 11e+7
Frequency	11	19	15	3
Percentage (%)	22.91667	39.58333	31.25	6.25
Cumulative (%)	22.91667	62.5	93.75	100
Mean	57268107			
Standard Deviation	33330673			
Median	54053789			
Mode	3128242			
Variance	1.11E+15			
R correlation	1	0.096768		
P matrix	1	0.512924		
	0.512924	1		

The return on assets ratio financial index (ROA) is a control variable indicating the size of the financial effectiveness index (calculated as the fraction of the net income over the assets' total value) showing the earnings/losses generated from the amount of the invested capital. The annual ROA results (mean= -12.3%, std=26%) is shown in Figure 35. Moreover, as shown in Table 23, the regression coefficient (R correlation) of these variables takes a positive value, with a confidence level greater than 95% since the p-value = 2.83E-14<0.05 corresponds to a non-significant correlation in R = 0.848 (see Figure 36).

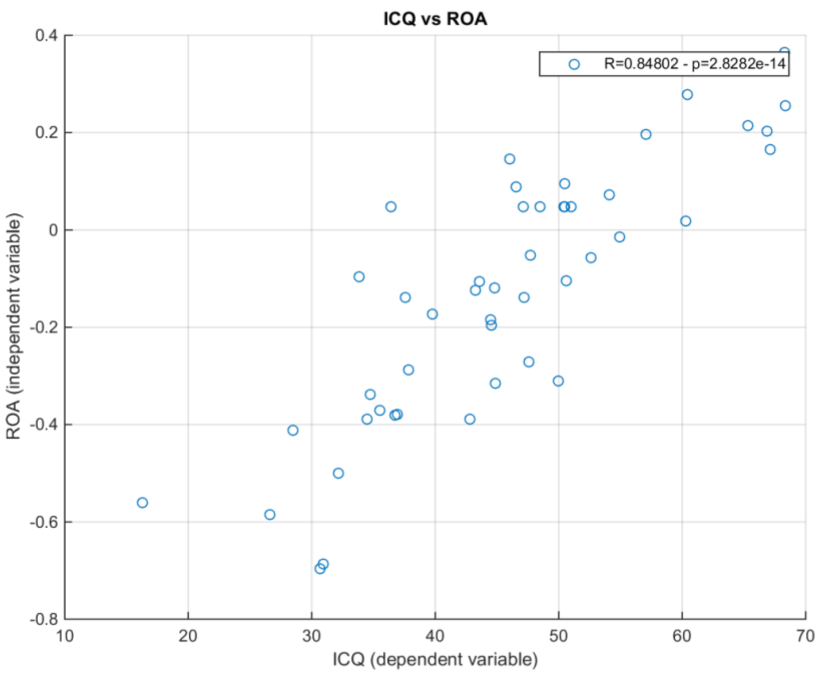

Figure 36. ICQ (dependent) Vs ROA (control) variable.

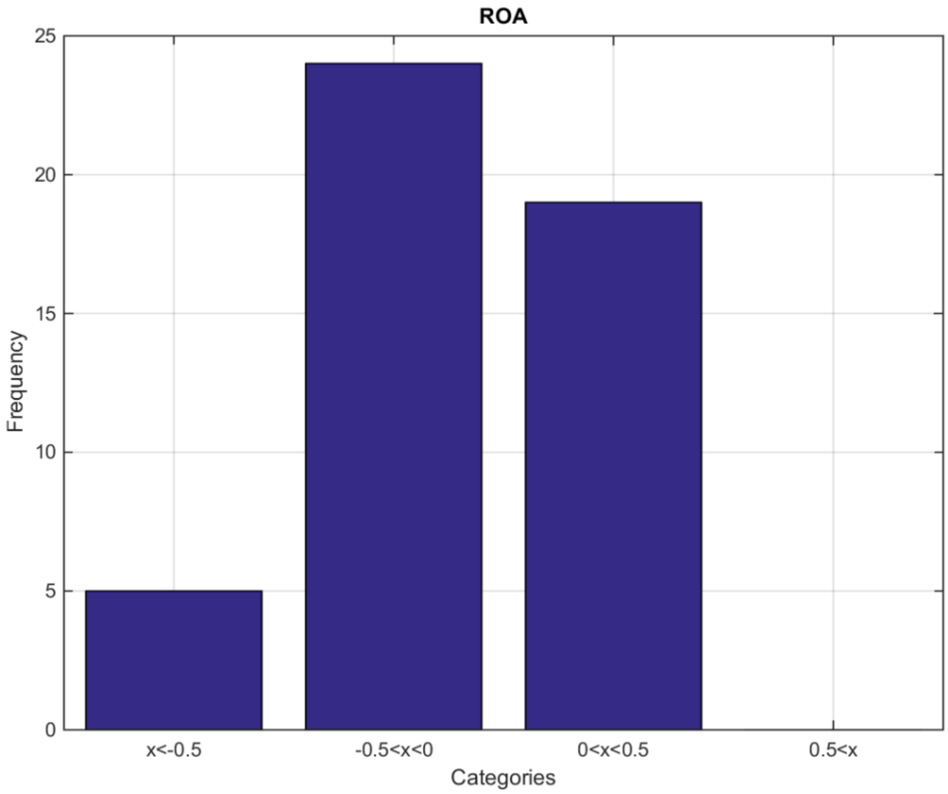

Figure 37. Return on assets ratio financial index (ROA).

Table 23. Descriptive statistics for ROA control variable.

ROA	x<-0.5	-0.5<x<0	0<x<0.5	x>0.5
Frequency	4	24	19	0
Percentage (%)	10.41667	50	39.58333	0
Cumulative (%)	10.41667	60.41667	100	100
Mean	-0.12364			
Standard Deviation	0.260496			
Median	-0.10458			
Mode	0.048038			
Variance	0.067858			
R correlation	1	0.848023		
P matrix	1	2.83E-14		
	2.83E-14	1		

The FIN_IND variable is a boolean control variable that equals one only if the firm belongs to any financial industry sector. The FIN_IND results (mean= 0.208, std=0.41) is shown in Figure 39. Moreover, as shown in Table 24, the regression coefficient (R correlation) of these variables takes a positive value, with a confidence level smaller than 95% since the p-value = 0.384>0.05 corresponds to a non-significant correlation in R = 0.128 (see Figure 38).

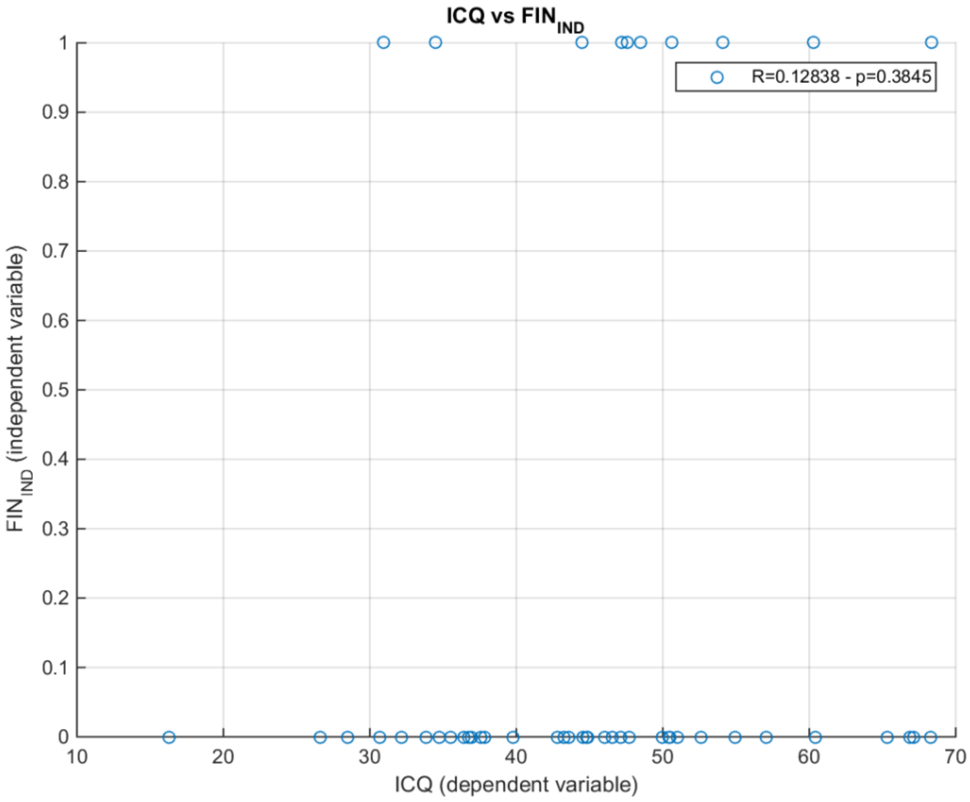

Figure 38. ICQ (dependent) Vs FIN_IND (control) variable.

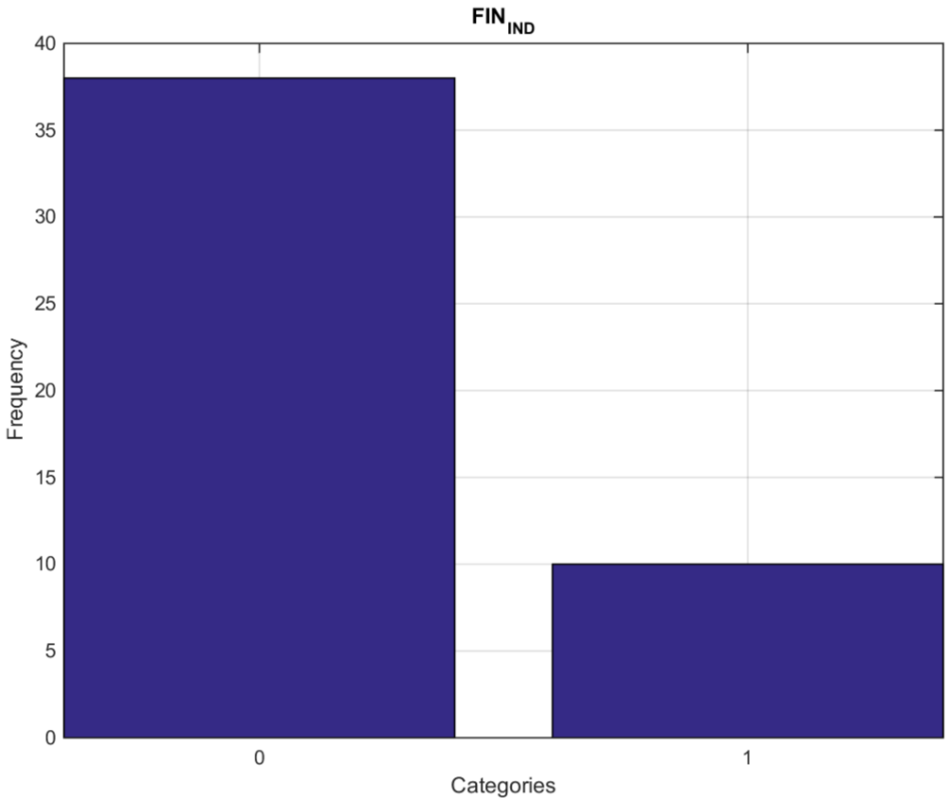

Figure 39. Boolean variable whether the firm belongs to any financial industry sector (FIN_IND).

Table 24. Descriptive statistics for FIN_IND control variable.

FIN_IND	0	1
Frequency	38	10
Percentage (%)	79.16667	20.83333
Cumulative (%)	79.16667	100
Mean	0.208333	
Standard Deviation	0.410414	
Median	0	
Mode	0	
Variance	0.16844	
R correlation	1	0.128384
P matrix	1	0.384503
	0.384503	1

6.2.3. The dependent variable

ICQ variable, is defined as ***the number of internal control deficiencies detected annually by chief executive auditors,*** considered as a representative index for the quality of internal control. Despite considering this variable as the independent one, it was also measured through the respective questionnaires sent to the committee members of the subject sample. The ICQ variable indicates the annual number of internal control deficiencies detected in each company. The ICQ results (mean= 45.768, std=11.6) is shown in Figure 40. Moreover, as shown in Table 24, the regression coefficient (R correlation) of these variables takes a positive value, with a confidence level greater than 95% since the p-value = 0 corresponds to a non-significant correlation in R = 1.

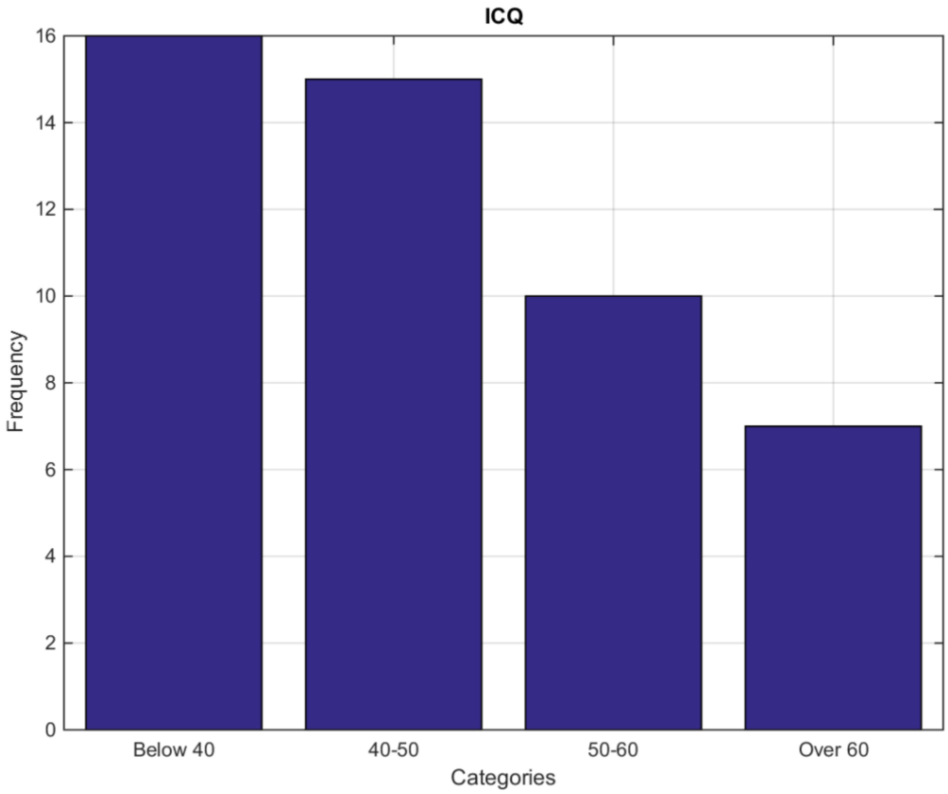

Figure 40. Internal control quality (ICQ).

Table 25. Descriptive statistics for ICQ dependent variable.

ICQ	Below 40	40-50	50-60	Over 60
Frequency	16	15	10	7
Percentage (%)	33.33333	31.25	20.83333	14.58333
Cumulative (%)	33.33333	64.58333	85.41667	100
Mean	45.76858			
Standard Deviation	11.60612			
Median	46.30675			
Mode	16.26738			
Variance	134.702			
R correlation	1	1		
P matrix	1	0		
	0	1		

6.3. Polynomial linear in the parameters regression results analysis

Since all variables have been processed accordingly in order to present a continuous behavior, several statistical polynomial regression models with various different maximum orders as well as total number of monomials were considered in order to detect the best statistical model choice. The generalized linear (in the parameters) regression model of polynomials has been formulated as follows:

$$ICQ \approx \widetilde{ICQ} = \beta_0 + \sum_{i=1}^{M} \beta_i \begin{pmatrix} (IAF_{OS})^{n1}(WKREL)^{n2}(STF_{COMP})^{n3}(IAF_{INV})^{n4}(QAS)^{n5} \\ (FUP_{DEF})^{n6}(FC_{XP})^{n7}(LN_{SLS})^{n8}(ROA)^{n9}(FIN_{IND})^{n10} \end{pmatrix}$$

Where "n1, n2, …, n10" are positive integer numbers, denoting the order of each independent variable contributing to the total order of the monomial as follows:

$$n1 + n2 + n3 + n4 + n5 + n6 + n7 + n8 + n9 + n10 \leq N$$

The order of each constituent independent variable in each monomial is determined randomly each time. The respective regression analysis for different maximum orders (N-values) is discussed within the following subsection respectively together with the respective training and validation errors. The total error for each respective case was averaged using the squared (L2) normalized formula as follows (where M is the total number of monomials):

$$e = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{ICQ - \widetilde{ICQ}}{ICQ} \right)^2$$

The methodology followed considers dividing the sampled items into two groups randomly which consider 75% of the sampled items as a training dataset for the regression model as well as the remaining 25% as the validation dataset where the trained model's accuracy and performance is being evaluated on a practically different/unknown dataset than the training one. Moreover, in order to avoid misguiding the training process due to the assumed sampled data noise and getting trapped into local minima, all independent variables were normalized to vary between [0, 10] and have the exact same scale of measure.

6.3.1. Training, Validation and Total Errors

The graphical representation of the training and validation errors for each different case of the selected number of monomials is shown in Section 8: Annex. As expected, the performance of the trained model over the training dataset is similar or better than the one achieved over the unknown validation dataset case. However, as a general conclusion the overall performance of the considered linear-in-the-parameters regression model is quite good, presenting total errors (both training and validation) close to zero. Moreover, as it was also expected, for smaller sizes of the polynomial the error of the trained model follows a decaying profile as the maximum order increases i.e. the performance improves as the maximum order increases. However, this dynamic disappears after a critical point of the number of monomials, where, as depicted, for both training and validation datasets, the fitting error presents a decaying behavior, for every selection of the maximum order from 1 to 15 of each monomial, up until the total number of monomials constituting the polynomial is around M=52 (see Figure 47). When the number of monomials increases beyond M=52, the performance of the trained linear (in the parameters) regression model for large values of the maximum monomial order is getting poorer and poorer.

Finally, as shown in Figure 41 and Figure 42, the problem of the maximum order and monomial number choice in multiple polynomial regression modelling techniques occurs in the considered dataset also. The relationship of the single dependent variable ICQ and the independent variables (see section 5.4) can adequately be emulated by a linear-in-the-parameters low-order polynomial since both validation and training dataset estimation errors are low in lower monomial number and maximum order choices. In these value ranges seem to be able to compensate the problem of generalization in regression problems, where the trained model is able to generalize adequately outside of its training set, presenting comparable estimation performance (Ostertagová, 2012).

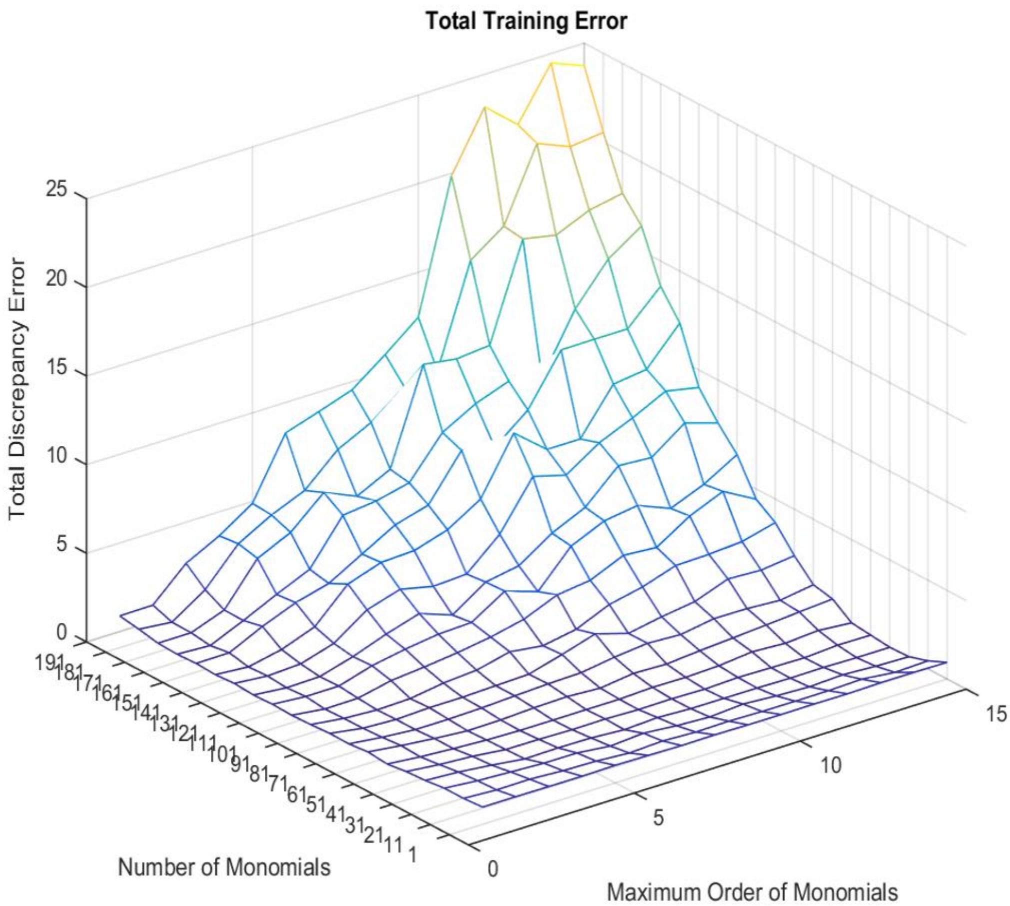

Figure 41. Total training errors (for different M and N values).

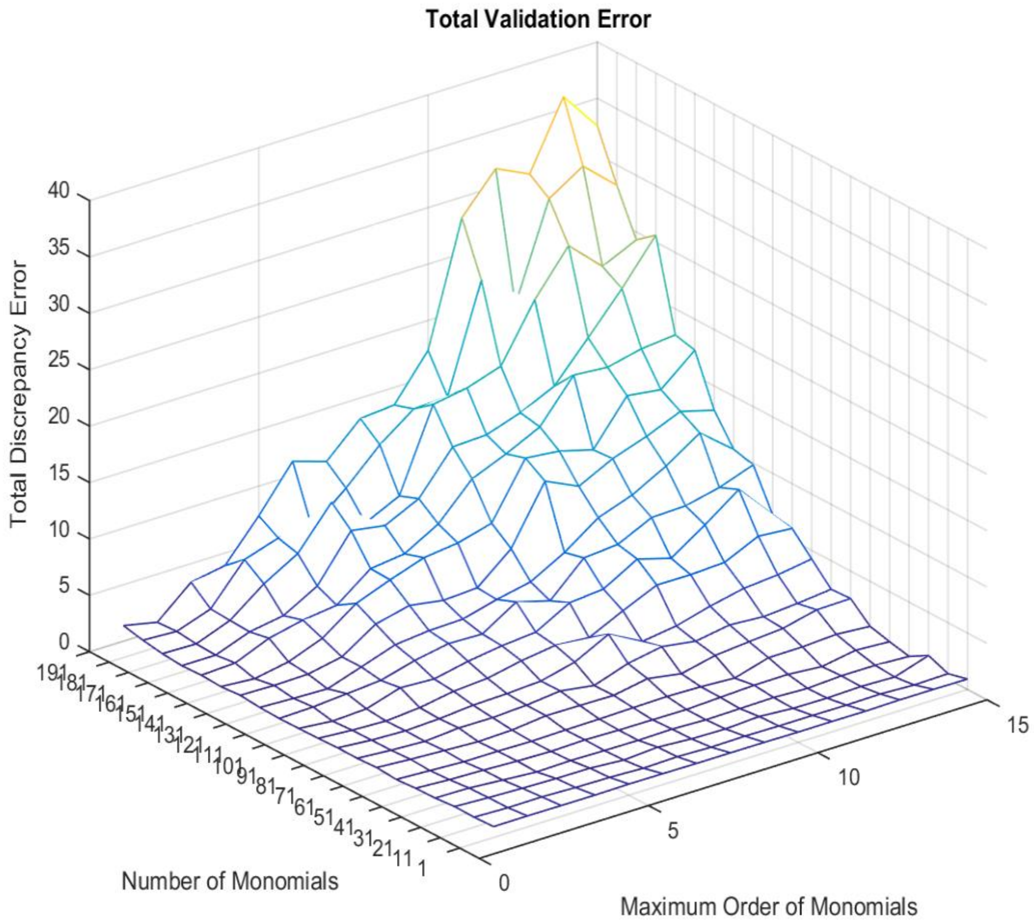

Figure 42. Total validation errors (for different M and N values).

6.3.2. Analytic Values for the trained linear regression model

This subsection presents the efficiency and effectiveness of the linear regression model version, considering maximum monomial order N=1 as well as the number of monomials equal to 11. The total estimation error, calculated using the formula discussed in section 6.3, both for the training as well as the validation datasets is equal to: Total Training Error + Total Validation Error = 0.94. The respective values of the theta vector coefficients are listed in Table 26 below.

Table 26. Theta Coefficients Values (M=11, N=1).

Coefficients-θ	Values
β1	0.010695
β2	0.06317
β3	-0.03872
β4	0.0402
β5	0.0367
β6	0.0556
β7	0.0872
β8	0.100695

β9	0.0911
β10	0.102361
β0 (constant term)	0.020653

While the regression elements' order, randomly generated, is shown in Table 27 below for each respective first-order monomial term.

Table 27. Monomial Orders (M=11, N=1).

	IAF_OS	WKREL	STF_COMP	IAF_INV	QAS	FUP_DEF	FC_XP	LN_SLS	ROA	FIN_IND	Total Order
β1	1	0	0	0	0	0	0	0	0	0	1
β2	0	1	0	0	0	0	0	0	0	0	1
β3	0	0	1	0	0	0	0	0	0	0	1
β4	0	0	0	1	0	0	0	0	0	0	1
β5	0	0	0	0	1	0	0	0	0	0	1
β6	0	0	0	0	0	1	0	0	0	0	1
β7	0	0	0	0	0	0	1	0	0	0	1
β8	0	0	0	0	0	0	0	1	0	0	1
β9	0	0	0	0	0	0	0	0	1	0	1
β10	0	0	0	0	0	0	0	0	0	1	1
β0	0	0	0	0	0	0	0	0	0	0	0

6.3.3. Analytic Values for the most-effective trained regression model

Under this subsection we select to present the performance (Total Training Error + Total Validation Error = 0.25) of the most effective polynomial regression model according to the already presented results considering M=31 (monomial number and a constant monomial scalar term) and N=12 (maximum monomial order). As a result, the total estimation error, with respect to the simplified linear case presented in subsection 6.3.2 is almost: (0.94-0.25)/0.94=73.4% better; or equivalently 3-4 times better, rendering the simplifying strategy considered for the linearized model version inadequately efficient when precision and more detailed estimation of the dependent ICQ variable are required. The respective values of the theta vector coefficients are listed in Table 28 below.

Table 28. Theta Coefficients Values (M=31, N=12).

Coefficients-θ	Values
β1	0.18087
β2	0.1957
β3	0.100695
β4	0.12946
β5	0.0705
β6	0.19609

β7	-0.17459
β8	0.14611
β9	0.15698
β10	0.18937
β11	0.19881
β12	0.16808
β13	0.197844
β14	0.18402
β15	0.18827
β16	0.16394
β17	0.06317
β18	0.16394
β19	0.18549
β20	0.18742
β21	-0.12547
β22	0.13971
β23	0.18734
β24	0.19415
β25	0.1864
β26	-0.17154
β27	-0.0912
β28	0.000845
β29	0.18767
β30	0.047346
β0 (constant term)	0.020653

As expected, the constant term value is exactly the same as in the linear case, discussed in sub section 6.2.2, denoting the standard offset in the ICQ measured values. While the regression elements' order, randomly generated, is shown in Table 29 below for each respective monomial term.

Table 29. Monomial Orders (M=31, N=12).

	IAF_OS	WKREL	STF_COMP	IAF_INV	QAS	FUP_DEF	FC_XP	LN_SLS	ROA	FIN_IND	Total Order
β1	0	0	0	0	0	4	0	0	0	0	4
β2	6	0	0	0	0	0	0	1	0	2	9
β3	0	1	0	0	0	0	2	0	10	1	14
β4	2	0	0	0	10	0	0	0	0	1	13
β5	1	9	0	0	0	0	0	0	0	0	10
β6	0	0	0	0	0	0	1	0	0	13	14
β7	0	0	6	8	0	0	0	0	0	0	14
β8	0	2	0	10	0	0	0	0	2	0	14
β9	0	0	0	7	1	0	0	0	1	0	9

β10	0	0	0	0	0	2	0	0	0	0	2
β11	0	3	0	0	0	0	1	0	3	0	7
β12	3	0	0	0	4	0	1	0	0	0	8
β13	0	0	0	0	0	0	0	0	1	0	1
β14	11	1	0	0	0	0	0	0	0	0	12
β15	0	0	0	0	6	0	0	0	0	0	6
β16	0	0	0	0	0	0	4	1	4	0	9
β17	0	0	0	0	4	0	0	0	0	0	4
β18	0	0	0	0	0	2	8	0	0	0	10
β19	8	0	0	4	0	0	0	1	1	0	14
β20	0	9	0	0	0	0	0	2	0	0	11
β21	3	2	2	0	0	0	1	0	2	0	10
β22	13	0	0	0	0	0	1	0	0	0	14
β23	0	0	0	10	0	0	1	0	0	0	11
β24	0	1	11	0	0	1	0	0	1	0	14
β25	0	4	0	0	0	0	0	10	0	0	14
β26	3	1	1	1	0	0	0	2	4	1	13
β27	1	4	4	0	1	1	0	0	1	0	12
β28	0	0	0	0	0	0	0	0	0	13	13
β29	0	1	0	5	3	0	0	0	1	1	11
β30	0	0	0	7	0	0	0	0	1	0	8
β0	0	0	0	0	0	0	0	0	0	0	0

As expected, the theta coefficient values (shown in Table 28 above), estimated using a multiple polynomial regression model, are reflecting the positive or negative relationship of each respective independent variable with the dependent one. For analysis simplicity reasons, consider, as an indicative example, the case of FUP_DEF independent variable; the monomial terms which consist only by FUP_DEF are monomial number M_1 and number M_{10}, where the order of the monomial is even (N_1=4 and N_{10}=2) while the respective values of the theta coefficients are both positive, indicating that FUP_DEF presents a positive relationship to ICQ while H6 hypothesis is true, confirming the respective conclusion derived from the descriptive statistics analysis in chapter 6.2.1.

Chapter 7: Conclusions and Future Work

This page was left blank intentionally.

7.1. Introduction

The current chapter consolidates briefly the results, the conclusions as well as the limitations considered within the presented methodology directly linked to proposed future work and steps.

7.2. Conclusions

Based on the evaluation analysis as well as the descriptive statistics, the research hypotheses defined in chapter 4.2 and discussed in chapter 6.2 were all true except H3. As a result, based on the results from 6.2, it could be reasonably assumed for the Greek enterprises case, with quite important statistical significance that:

- *H1. There is a positive association between IAF organizational status and ICQ.*
- *H2. There is a positive association between the audit committee's involvement in reviewing the IAF execution and ICQ.*
- *H3. There is no positive association between IAF competence and ICQ.*
- *H4. Allocating greater resources for IAF leads to less severe internal control weaknesses.*
- *H5. There is a positive association between internal audit quality assurance and ICQ.*
- *H6. There is a positive association between the existence of follow-up process and internal control quality.*

The aforementioned statistical problem was formulated considering only Internal Control Quality (ICQ) as the dependent variable, 6 independent and 4 control variables. Recent advances in ERP operational ICT systems, event and data logging mechanisms, efficient data manipulation and computational devices, enables generating and collecting huge amounts of corporate data. This type of data may usually contain crucial information for corporate processes malfunctions and fraudulent behaviors which could affect the operational quality and capacity of the whole company in the short or long term. Therefore, for example, more than 6 independent and 4 control variables could be imposed in the formulation of the aforementioned correlation/regression problem in order to achieve better ICQ estimation. It is more than evident that as the scale of the problem increases the complexity of the descriptive statistics analysis increases in an analogous manner. The main problem that the current study tries to tackle is the exploitation/utilization strategies for big amounts of data. Motivated by recent literature and advances in statistical mathematics, image processing, artificial intelligence, the current study explores the potential of adopting similar ICT tools for internal auditing assessment purposes.

The study adopts a multiple polynomial regression model (see chapter 5.4.2), explores the best (the most efficient one in terms of approximation error) option among different choices of monomial number and order and concludes to data analytics tools and well-established regression methodologies for model structure exploration and estimation, lead to similar results, as well as reconfirm the exact same 6 research hypotheses (as the ones established using conventional descriptive statistics analysis).

A performance comparison between the best resulted polynomial regression model, considering M=31 monomials and N=12 maximum order (see chapter 6.3.3), presents over 3 times better estimation performance (3 times smaller total approximation error) both for the training (known) and validation (unknown) dataset cases, as compared with the most common simplified case of a linear regression model (see chapter 6.3.2).

The power of data analytics and computerized programming tools is evident when, despite the fact that the sample size was quite small to extract safe conclusions, the calculated regression model is usable for other Greek enterprises' cases in order to predict the respective ICQ levels in a straightforward manner (considering similar macro-economic exogenous conditions such as political stability, environmental/weather, societal, technological).

Conclusively, what has to be underlined is the big importance of data analytics and information mining in internal control as a tool for effective decision making and consequently, compensation of the corporate risks as well as corporate governance.

7.3. Limitations & Future Steps

The current study presents implementation as well as case-study related limitations. The main limitations of the study derive from the research conduction short period and time constraints, allowing only for a 5-month work to be carried out. As a consequence, the sample as well as the time allowed to the sampled auditors opinion to fill the questionnaire was also limited, narrowing the timeframe for revisions or reconsiderations. Moreover, international literature studies for the Greek internal auditing quality are quite scarce, where similar studies could be utilized as a benchmark case for the current or other future works.

Therefore, since empirical research together with data analysis and information extraction in such large data sets, are considered, the allocation of larger periods of time as well as more dedicated personal interviews (instead of impersonal questionnaires) could assist on extracting much more concrete conclusions. Finally, practical limitations related to the parametrization of the regression modelling problem, such as data pre-processing (preparation), data structuring, regression modelling order, number of monomials, monomial order generation strategy, etc. could be considered also as potential barriers that should be studied further. There may be different or additional variables affecting the dependent ICQ variable values which render the choice of different or additional variables (such as the CEO background and characteristics, the technological information systems utilized, the fraudulent behavior frequency and the legislation framework strictness, etc.) for formulating the ICQ levels as an important aspect for future studies. Moreover, the different range of exploring the maximum monomial order or the number of monomials could also present significant research interest in future studies.

This page was left blank intentionally.

Chapter 8: Annex

This page was left blank intentionally.

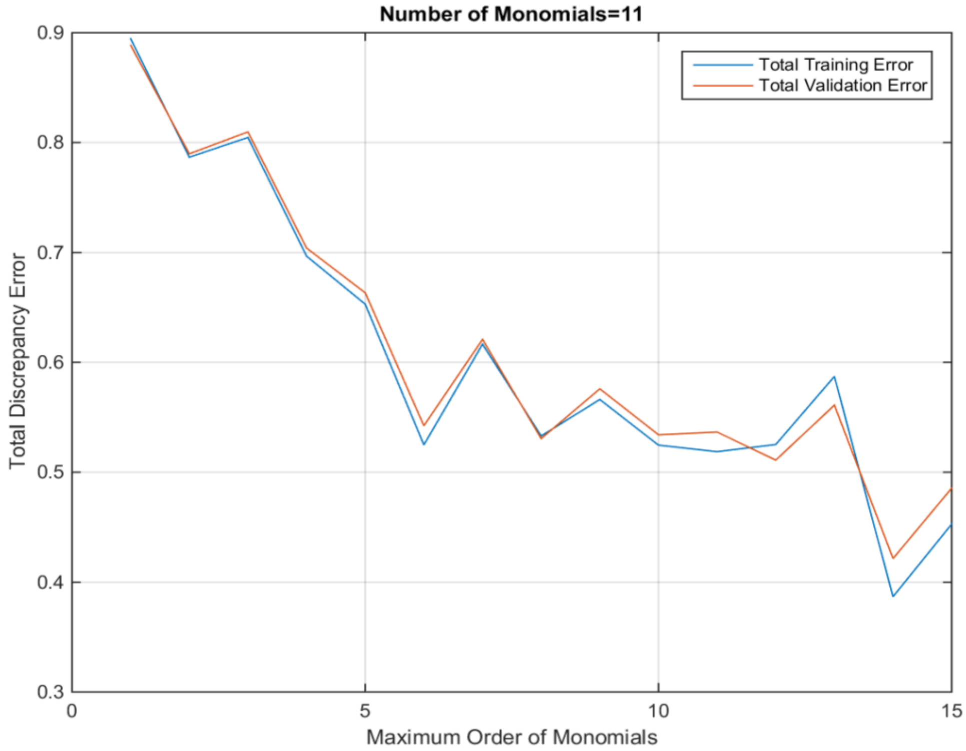

Figure 43. Training and Validation Dataset Error for number of monomials (M=10+1 constant term) and varying monomial orders (N=1:1:15).

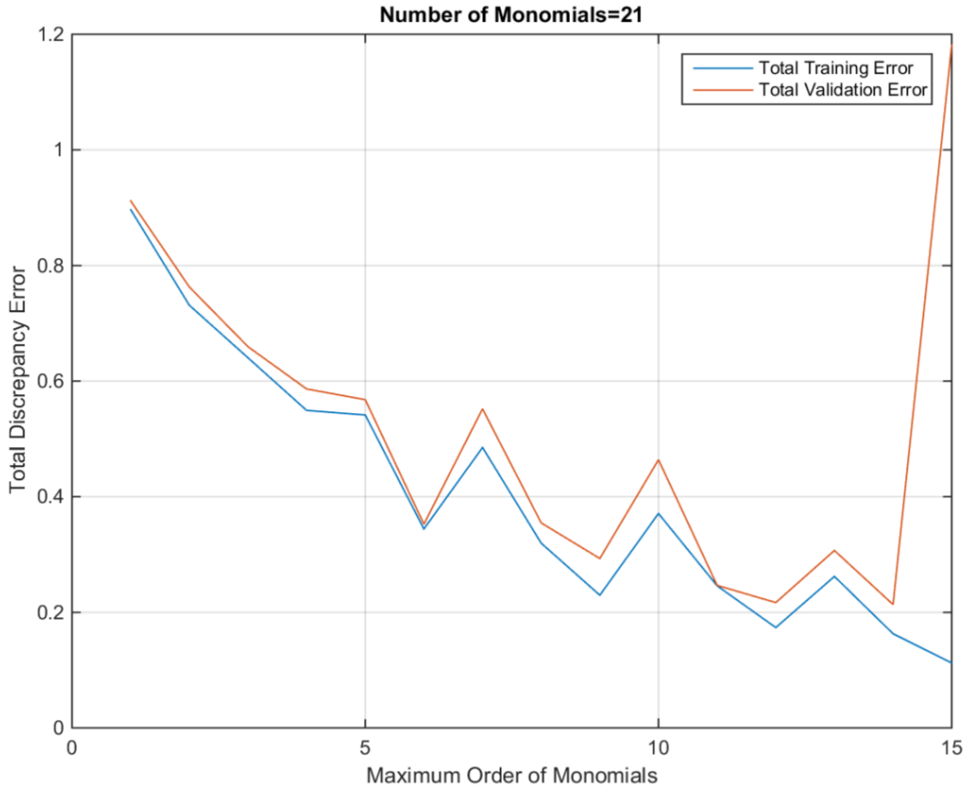

Figure 44. Training and Validation Dataset Error for number of monomials (M=20+1 constant term) and varying monomial orders (N=1:1:15).

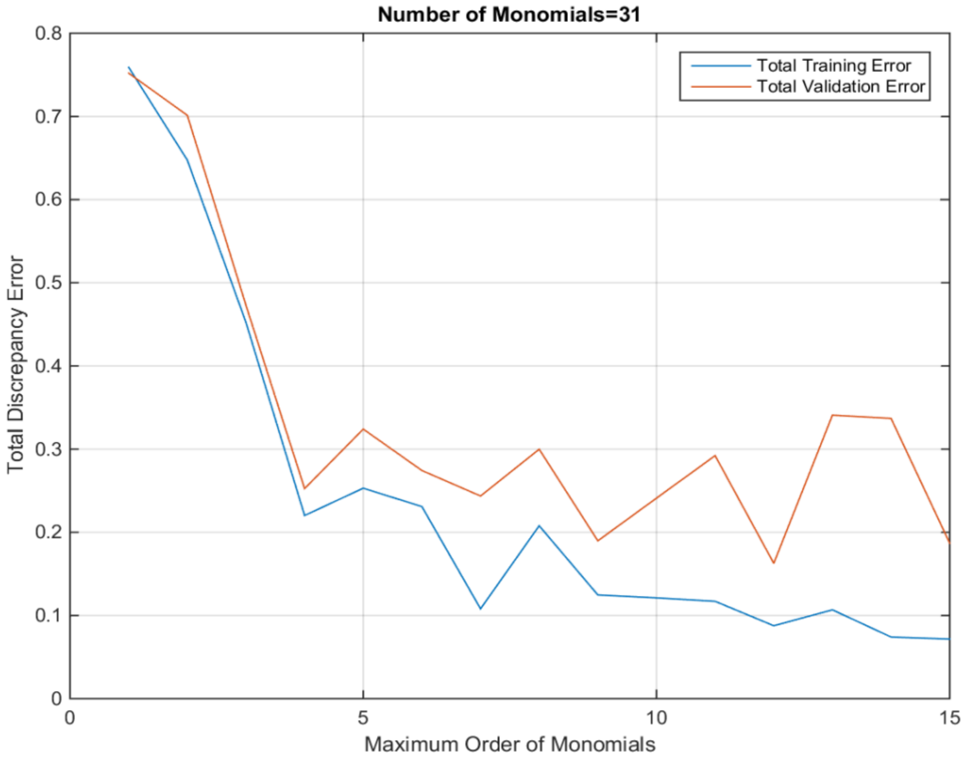

Figure 45. Training and Validation Dataset Error for number of monomials (M=30+1 constant term) and varying monomial orders (N=1:1:15).

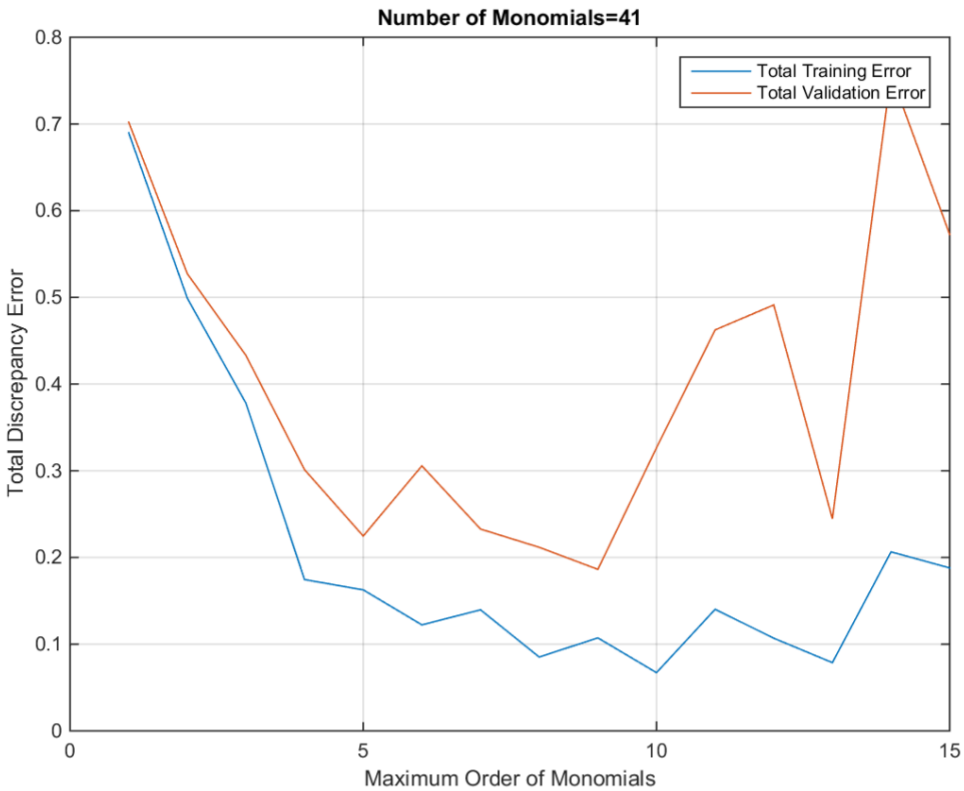

Figure 46. Training and Validation Dataset Error for number of monomials (M=40+1 constant term) and varying monomial orders (N=1:1:15).

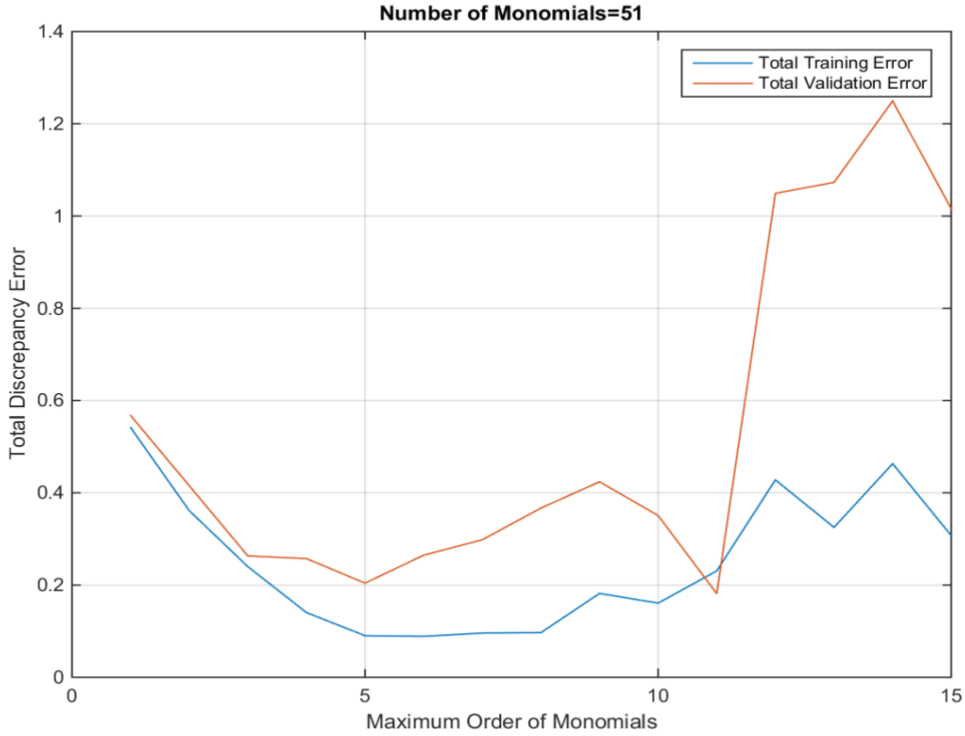

Figure 47. Training and Validation Dataset Error for number of monomials (M=50+1 constant term) and varying monomial orders (N=1:1:15).

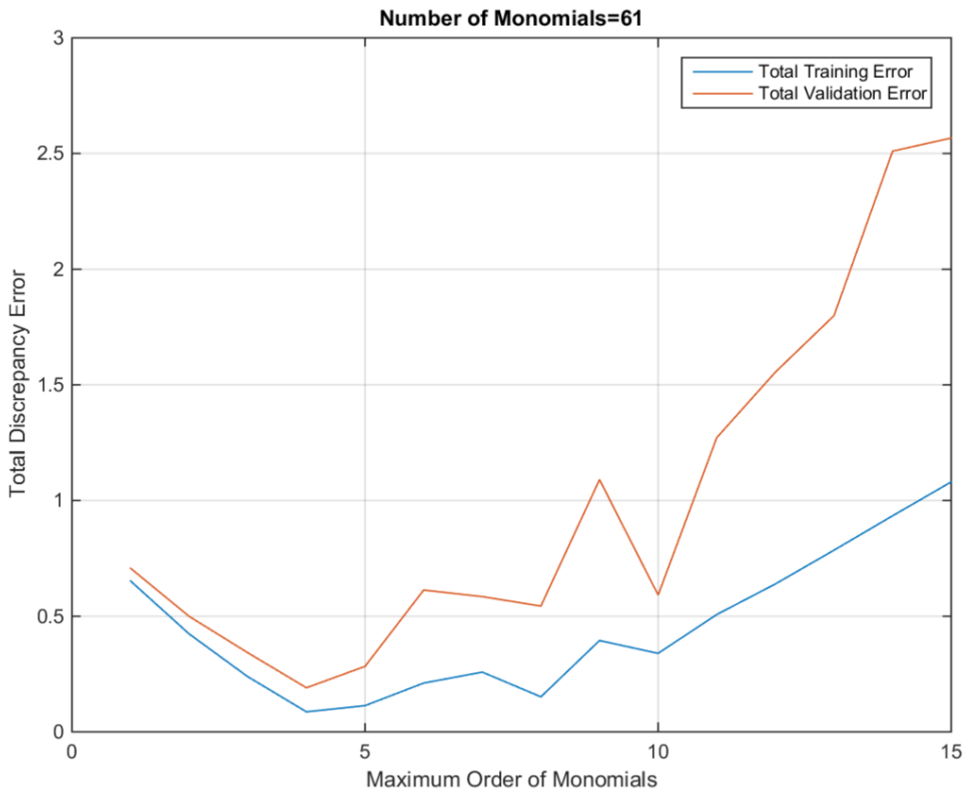

Figure 48. Training and Validation Dataset Error for number of monomials (M=60+1 constant term) and varying monomial orders (N=1:1:15).

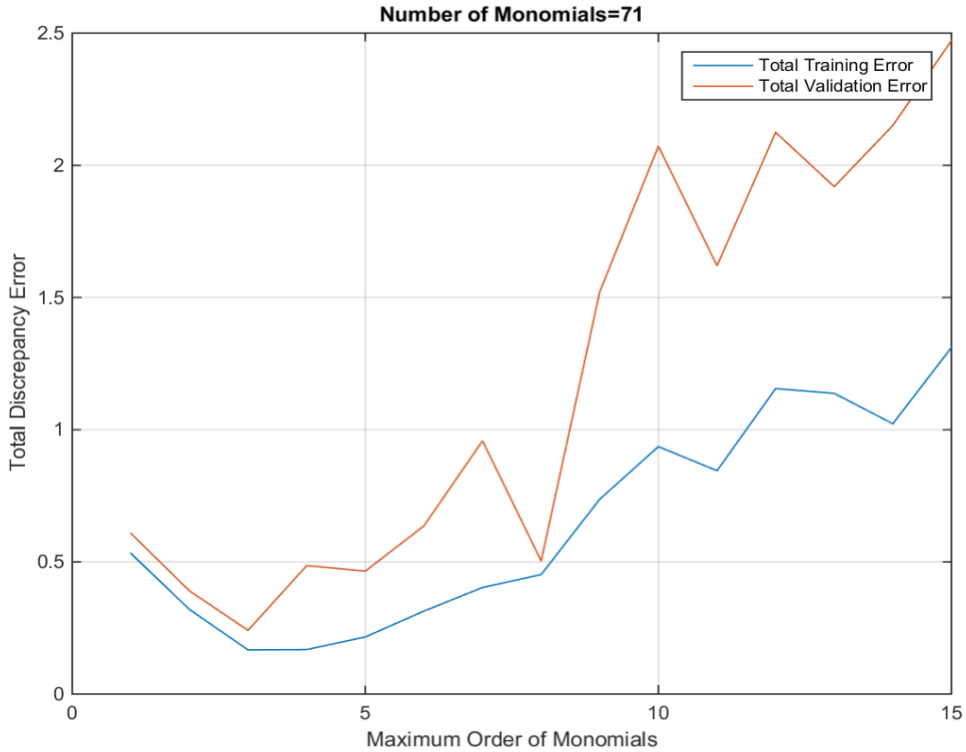

Figure 49. Training and Validation Dataset Error for number of monomials (M=70+1 constant term) and varying monomial orders (N=1:1:15).

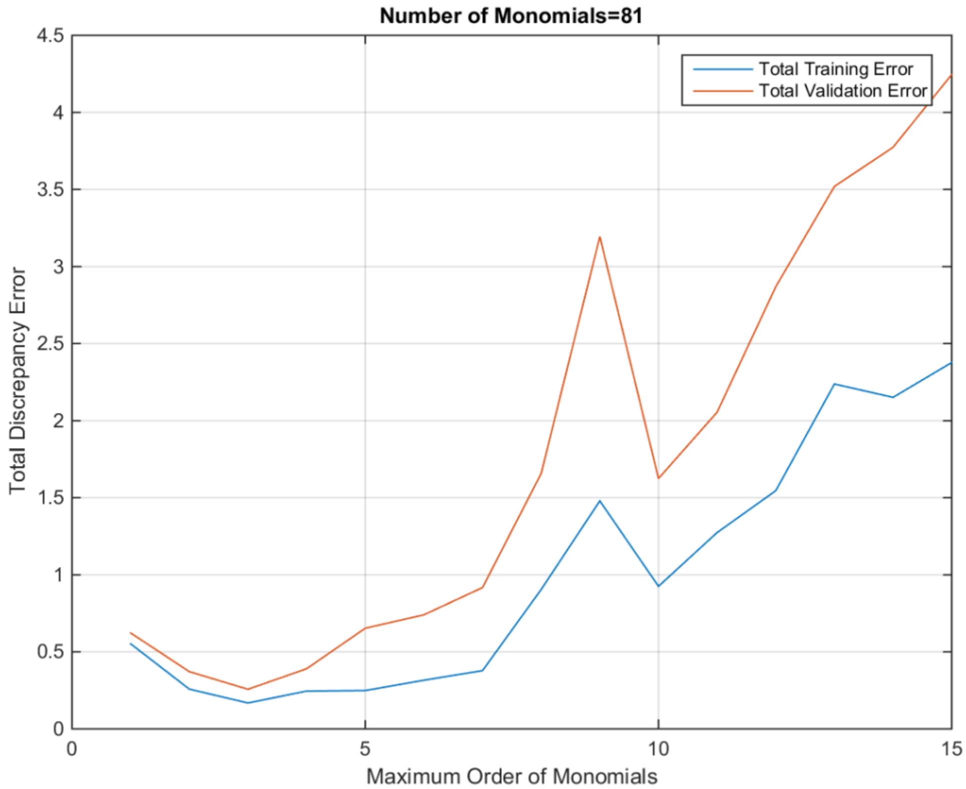

Figure 50. Training and Validation Dataset Error for number of monomials (M=80+1 constant term) and varying monomial orders (N=1:1:15).

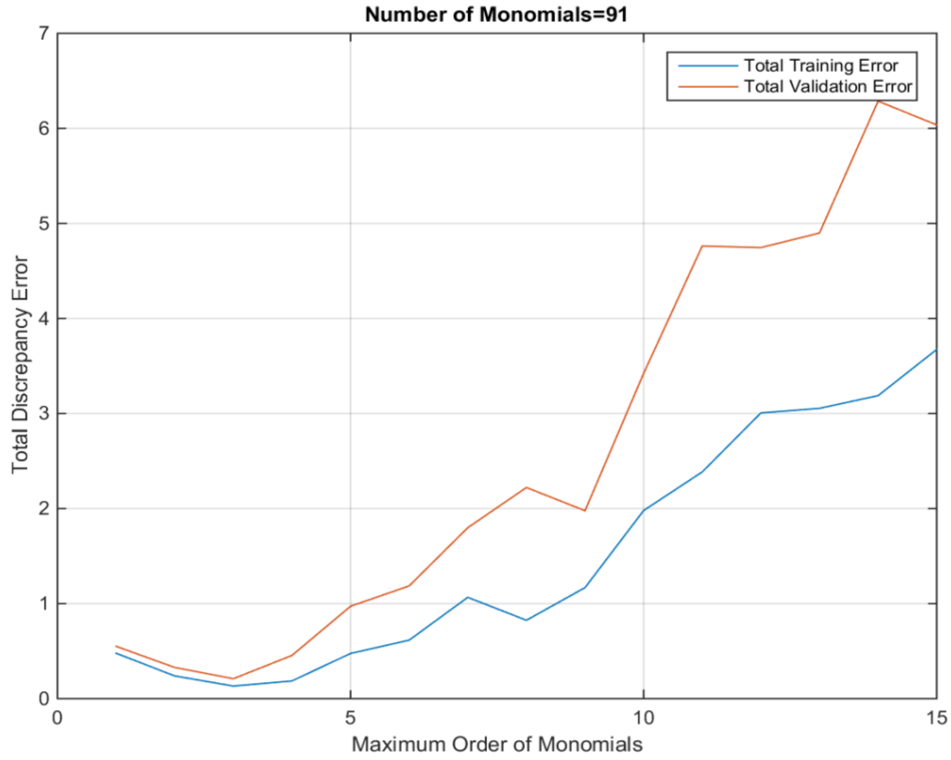

Figure 51. Training and Validation Dataset Error for number of monomials (M=90+1 constant term) and varying monomial orders (N=1:1:15).

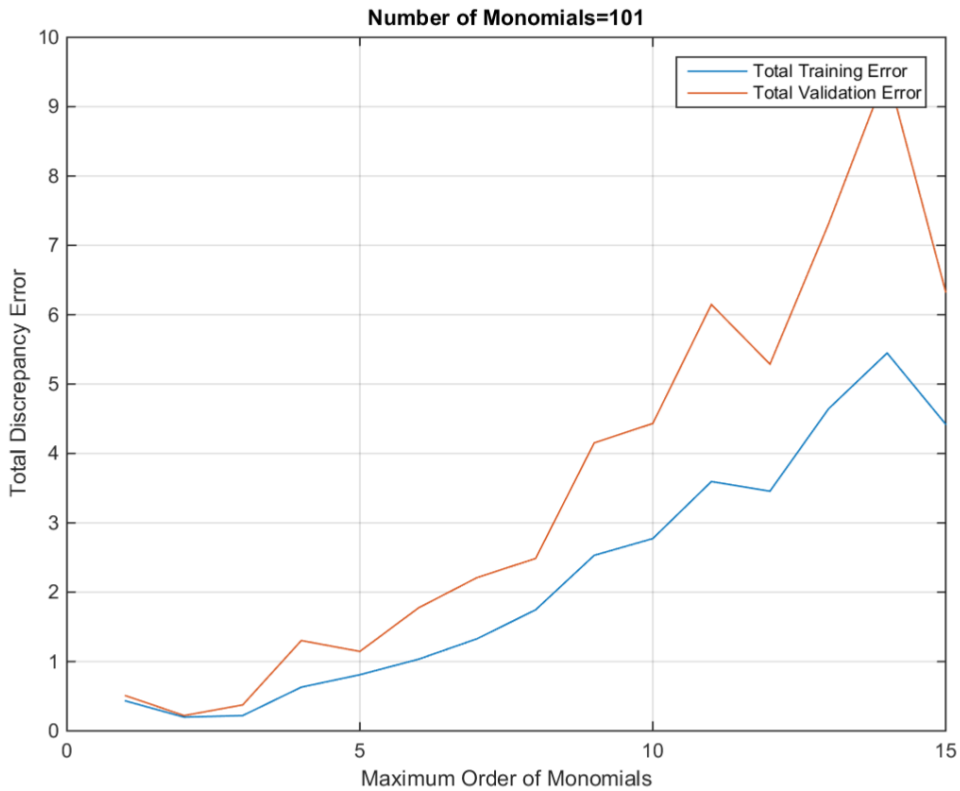

Figure 52. Training and Validation Dataset Error for number of monomials (M=100+1 constant term) and varying monomial orders (N=1:1:15).

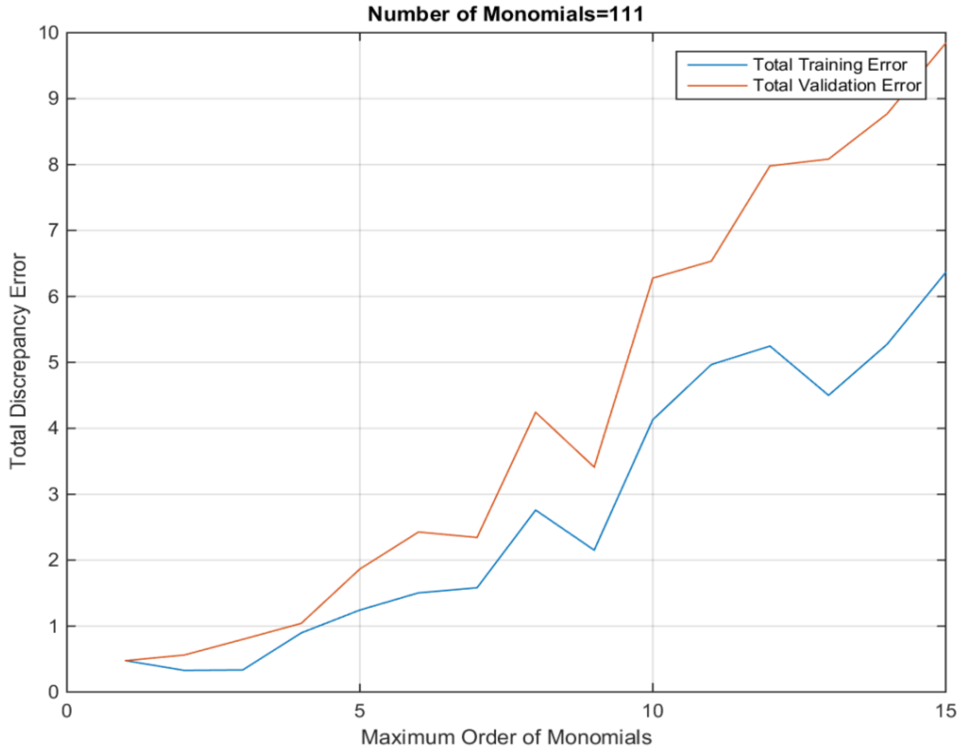

Figure 53. Training and Validation Dataset Error for number of monomials (M=110+1 constant term) and varying monomial orders (N=1:1:15).

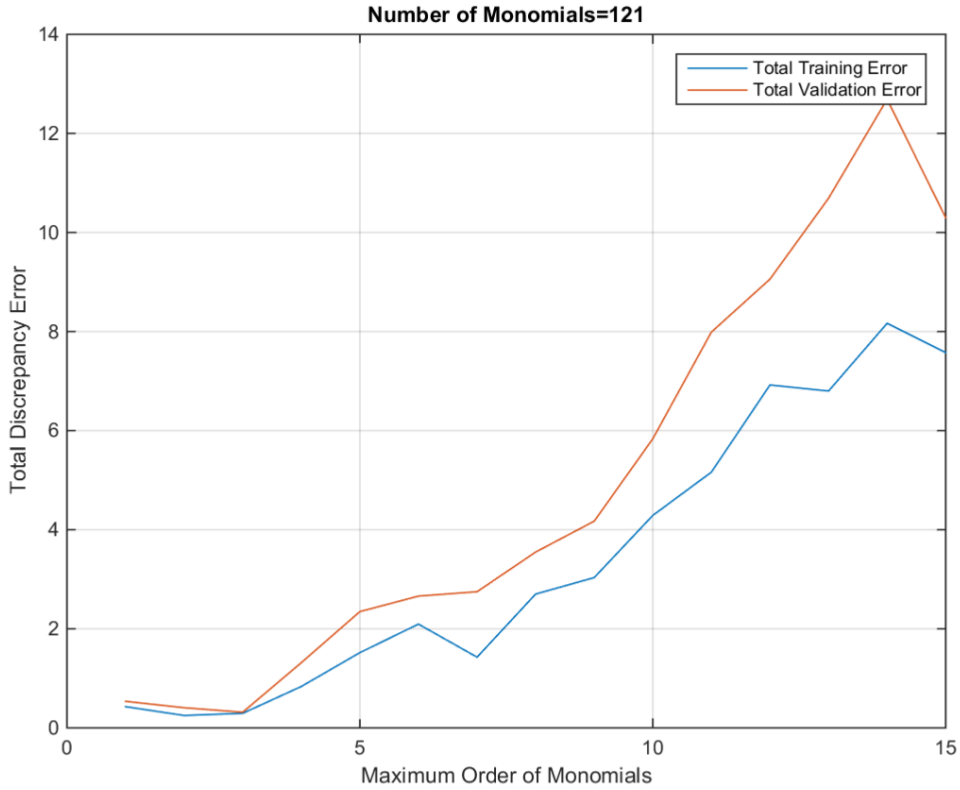

Figure 54. Training and Validation Dataset Error for number of monomials (M=120+1 constant term) and varying monomial orders (N=1:1:15).

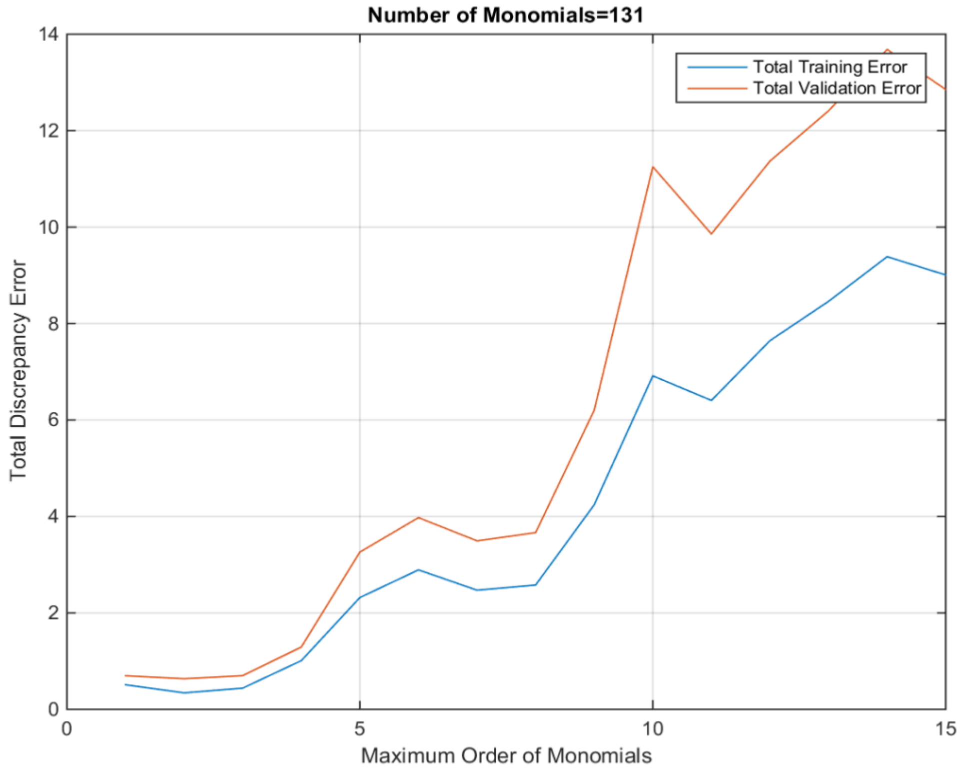

Figure 55. Training and Validation Dataset Error for number of monomials (M=130+1 constant term) and varying monomial orders (N=1:1:15).

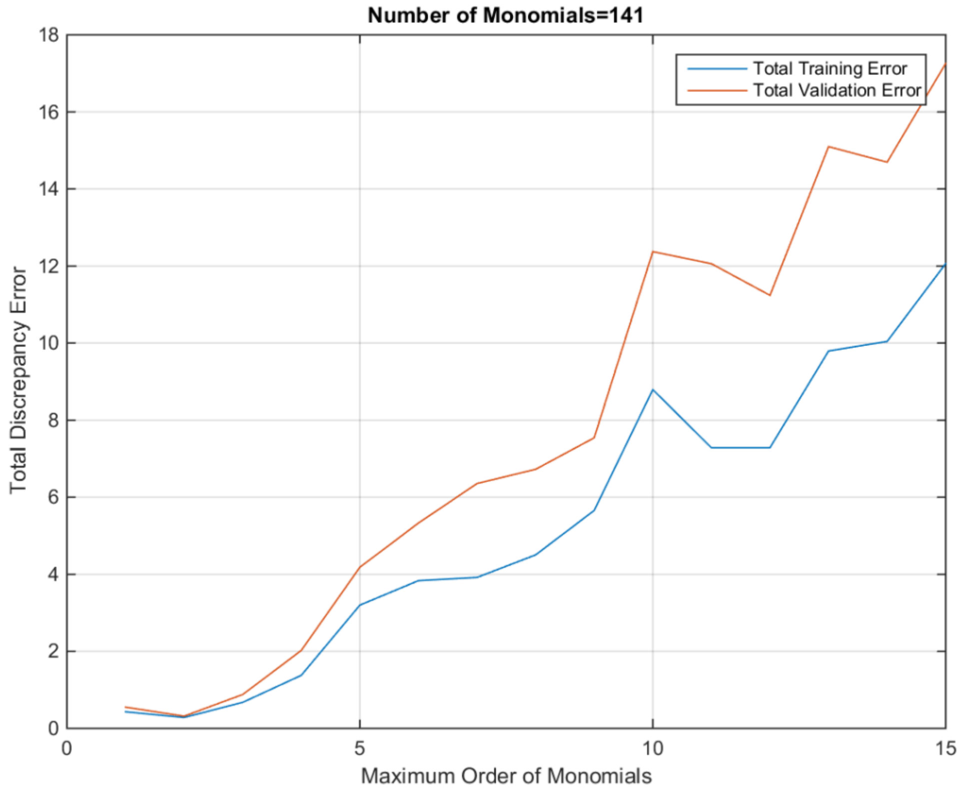

Figure 56. Training and Validation Dataset Error for number of monomials (M=140+1 constant term) and varying monomial orders (N=1:1:15).

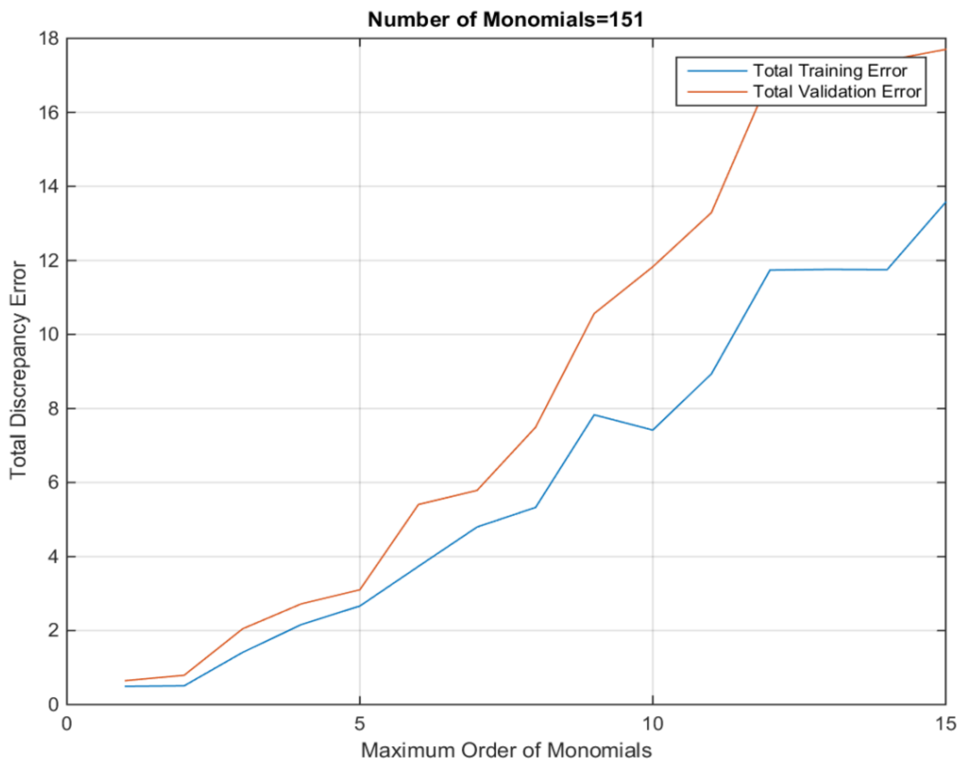

Figure 57. Training and Validation Dataset Error for number of monomials (M=150+1 constant term) and varying monomial orders (N=1:1:15).

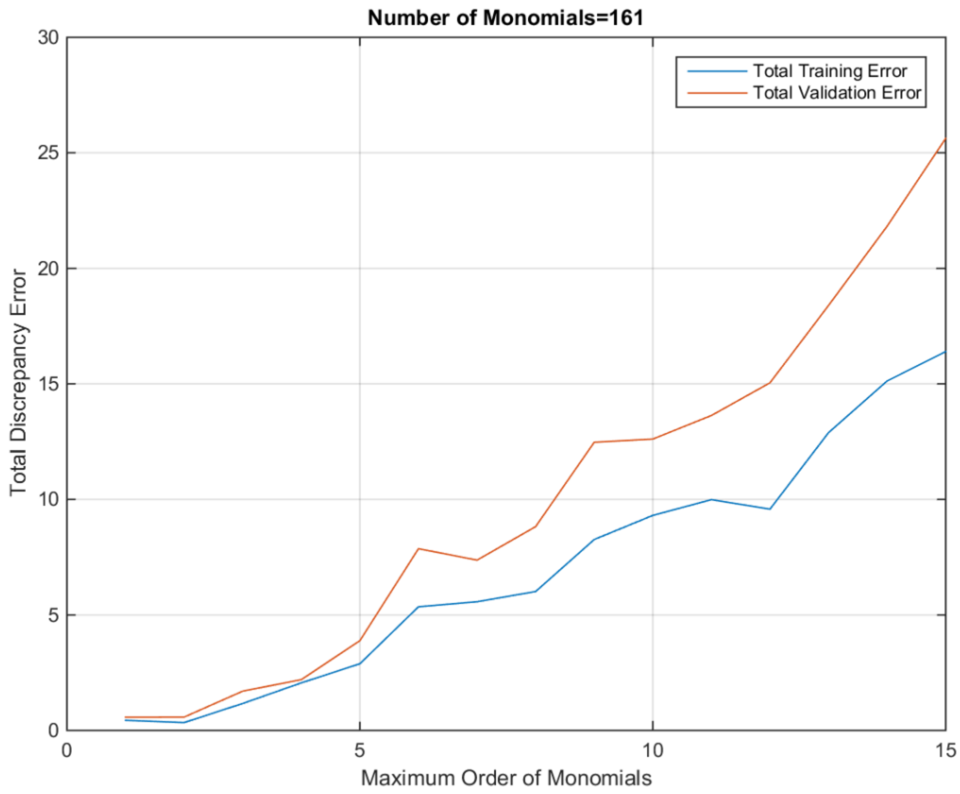

Figure 58. Training and Validation Dataset Error for number of monomials (M=160+1 constant term) and varying monomial orders (N=1:1:15).

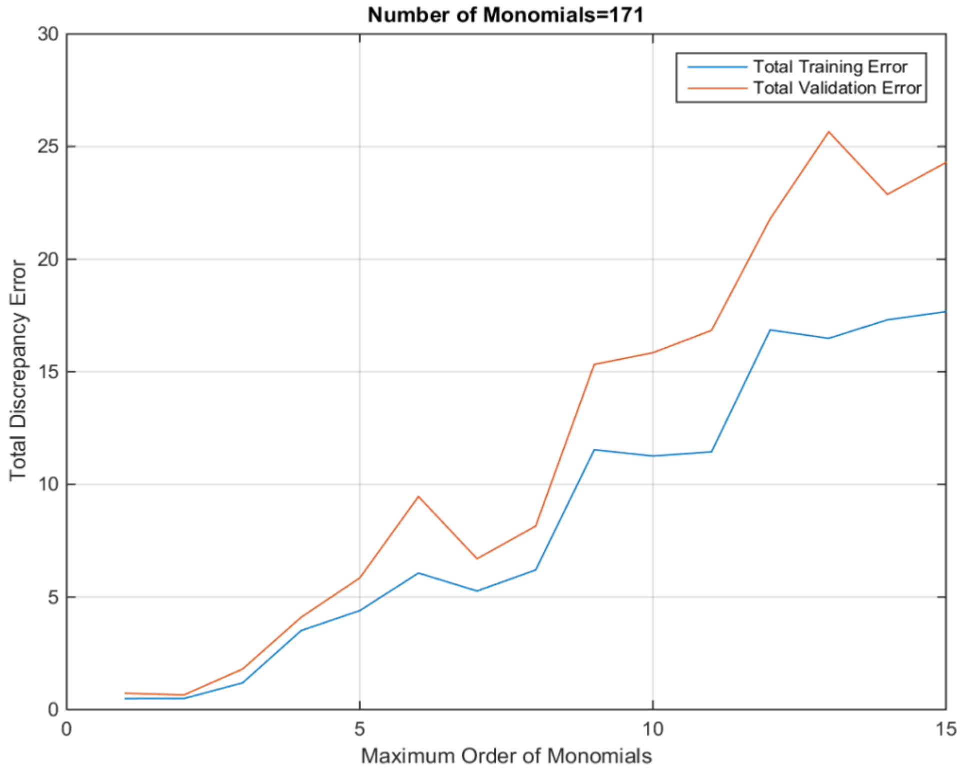

Figure 59. Training and Validation Dataset Error for number of monomials (M=170+1 constant term) and varying monomial orders (N=1:1:15).

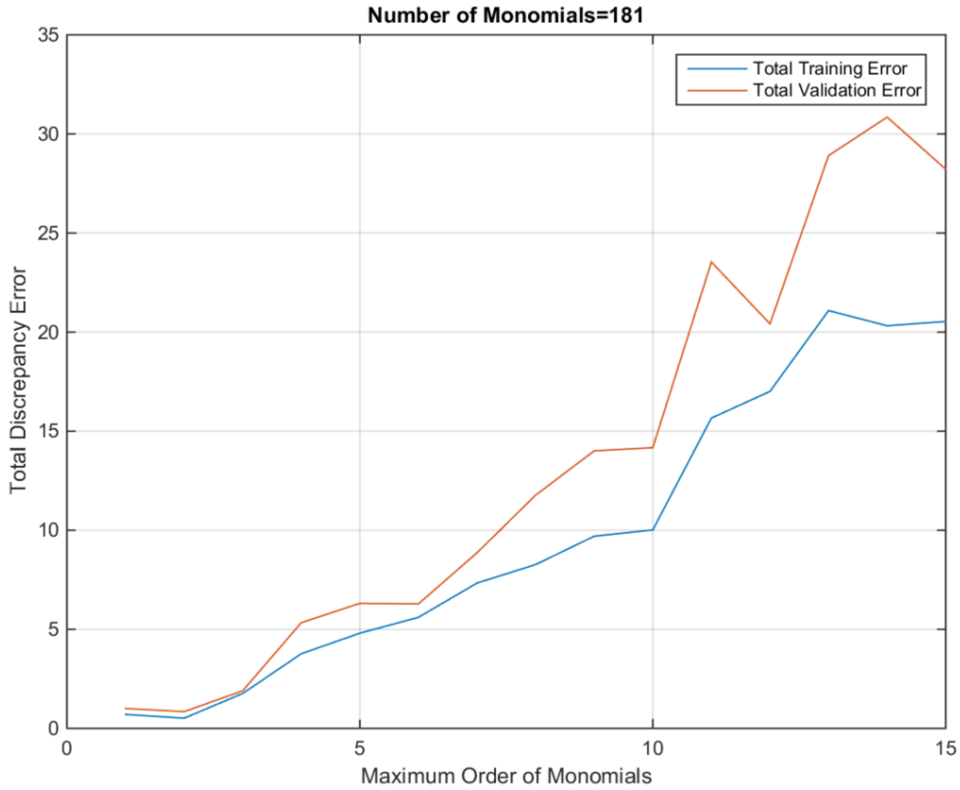

Figure 60. Training and Validation Dataset Error for number of monomials (M=180+1 constant term) and varying monomial orders (N=1:1:15).

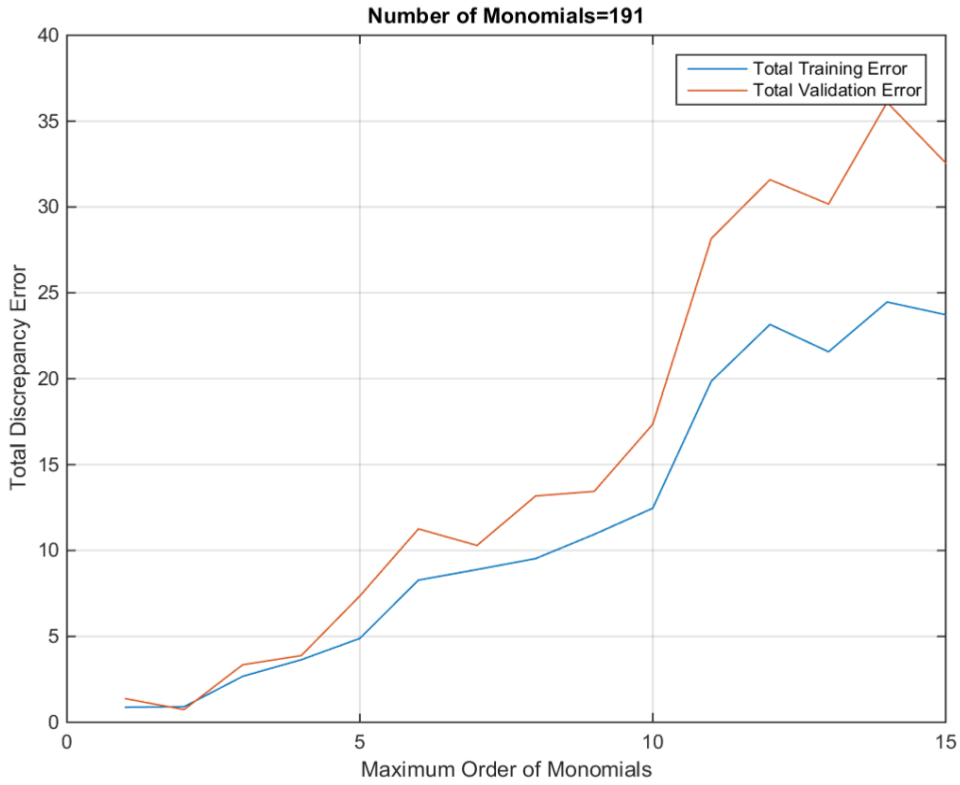

Figure 61. Training and Validation Dataset Error for number of monomials (M=190+1 constant term) and varying monomial orders (N=1:1:15).

Chapter 9: Bibliography

This page was left blank intentionally.

References

Abbott, L., Parker, S., & Peters, G. (2010). Serving two masters: the association between audit committee internal audit oversight and internal audit activities. *Accounting Horizons, 24*(1), 1-2. doi:https://doi.org/10.2308/acch.2010.24.1.1

Alhajeri, M. (2017). Factors associated with the size of internal audit functions: evidence from Kuwait. *Managerial Auditing Journal, 32*(1), 75-89.

Al-Khaddash, H., Al Nawas, R., & Ramadan, A. (2013, September). Factors affecting the quality of auditing: The case of Jordanian commercial banks. *International Journal of Business and Social Science, 4*(11). Retrieved from https://pdfs.semanticscholar.org/ff2e/359ee70eb3cbcb5854184aa4756ba12a473a.pdf

Alvin, A., & Loebbecke, J. (1999). *Auditing: An Integrated Approach* (8th ed.). Prentice Hall College Div.

Alzeban, A., & Sawan, N. (2015). The impact of audit committee characteristics on the implementation of internal audit recommendations. *Journal of International Accounting, Auditing and Taxation, 24*(5), 61-71.

Arena, M., & Azzone, G. (2009). Identifying organizational drivers of internal audit effectiveness. *International Journal of Auditing, 13*(1), 43-57. doi:https://doi.org/10.1111/j.1099-1123.2008.00392.x

Bame-Aldred, C., Brandon, D., Messier, W. J., Rittenberg, L., & Stefaniak, C. (2013). A summary of research on external auditor reliance on the internal audit function. *Auditing: A Journal of Practice & Theory, 32*(1), 251-286. doi:https://doi.org/10.2308/ajpt-50342

Barroso-Castro, C., Villegas-Periñan, M. D., & Casillas-Bueno, J. C. (2016). How boards' internal and external social capital interact to affect firm performance. *Strategic Organization, 14*(1), 6-31.

Becker, C. L., DeFond, M. L., Jiambalvo, J., & Subramanyam, K. R. (1998). The effect of audit quality on earnings management. *Contemporary accounting research, 15*(1), 1-24.

Bedard, J., & Graham, L. (2011). Detection and severity classifications of sarbanes-oxley section 404 internal control deficiencies. *International Journal of Auditing, 86*(3), 825-855.

Berson, S., & Thearling, K. (1999). *Building Data Mining Applications for CRM*. New York, USA: Mc-Graw Hill Companies Inc.

Burton, F. G., Emett, A. S., Simon, A. C., & Wood, A. D. (2012). Corporate Managers' Reliance on Internal Auditor Recommendations. *AUDITING: A Journal of Practice & Theory, 31*(2), 151-166. doi:https://doi.org/10.2308/ajpt-10234

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0*. Retrieved from CRISPMWP-1104-1C.qxd: https://the-modeling-agency.com/crisp-dm.pdf

Data mining. (2019, March 15). Retrieved from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/w/index.php?title=Data_mining&oldid=886766563

De Silva Lokuwaduge, C., & Armstrong, A. (2015). The impact of governance on the performance of the higher education sector in Australia. *Educational Management Administration & Leadership, 43*(5), 811-827.

De Zwaan, L., Stewart, J., & Subramaniam. (2011). Internal audit involvement in enterprise risk management. *Managerial Auditing Journal, 26*(7), 586–604.

Endaya, K., & Hanefah, M. (2016). Internal auditor characteristics, internal audit effectiveness, and moderating effect of senior management. *Journal of Economic and Administrative Sciences, 32*(2), 160-176.

Fadzil, F., Haron, H., & Jantan, M. (2005). Internal auditing practices and internal control system. *Managerial Auditing Journal, 20*(8), 844-866.

Fraser, I., & Henry, W. (2007). Embedding risk management: structures and approaches. *Managerial Auditing Journal, 22*(4), 392-409.

Goodwin-Stewart, J., & Kent, P. (2006). The use of internal audit by Australian companies. *Managerial Auditing Journal, 21*(1), 81-101.

Gramling, A., & M., M. P. (2006). Internal Auditing's Role in ERM. *The Internal Auditor, 63*(2), 52-58. Retrieved from http://www.theiia.org/intAuditor/featurearticles/2006/april/internal-auditings-role-in-erm/

Han, K., Pei, J., & Micheline, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* Cambridge, Massachusetts: MIT Press.

Hope, O. K., Thomas, W. B., & Vyas, D. (2013). Financial reporting quality of US private and public firms. *The Accounting Review, 88*(5), 1715-1742.

IIA, I. o. (2004, May 26). INTERNAL AUDITING S ROLE IN SECTIONS 302 AND 404 OF THE U.S. SARBANES-OXLEY ACT OF 2002. Altamonte Springs, Florida, USA. Retrieved from https://na.theiia.org/standards-guidance/Public%20Documents/Act%20Internal_Auditings_Role_in_Sections_302__404__FINAL.pdf

IIA, I. o. (2016). *INTERNATIONAL STANDARDS FOR THE PROFESSIONAL PRACTICE OF INTERNAL AUDITING.* Lake Mary, FL: IIA, Institute of Internal Auditors. Retrieved from https://global.theiia.org/standards-guidance/Public%20Documents/IPPF-Standards-2017.pdf

IMF, I. M. (2018, October). *World Economic Outlook Database.* Retrieved from International Monetary Fund: https://www.imf.org/external/pubs/ft/weo/2018/02/weodata/weoselgr.aspx

Internal audit. (2019, March 15). Retrieved from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/w/index.php?title=Internal_audit&oldid=886395596

Johl, S., Subramaniam, N., & Cooper, B. (2013). Internal audit function, board quality and financial reporting quality: evidence from Malaysia. *Managerial Auditing Journal, 28*(9), 780-814.

Khlif, H., & Samaha, K. (2014). Internal control quality, Egyptian standards on auditing and external audit delays: evidence from the Egyptian stock exchange. *International Journal of Auditing, 18*(2), 139-154.

Khlif, H., & Samaha, K. (2016). Audit committee activity and internal control quality in Egypt: does external auditor's size matter? *Managerial Auditing Journal, 31*(3), 269-289.

Kinney Jr., W. R., Palmrose, Z. V., & Scholz, S. (2004). Auditor independence, non-audit services, and restatements: Was the US government right? *Journal of Accounting Research, 42*(3), 561-588.

Lenz, R., Sarens, G., & D'Silva, K. (2014). Probing the discriminatory power of characteristics of internal audit functions: sorting the wheat from the Chaff. *International Journal of Auditing, 18*(2), 126-138.

Lin, S., Pizzini, M., Vargus, M., & Bardhan, I. (2011). The role of the internal audit function in the disclosure of material weaknesses. *The Accounting Review, 86*(1), 287-323.

Lin, Y., Wang, Y., Chiou, J., & Huang, H. (2014). CEO characteristics and internal control quality. *Corporate Governance: An International Review, 22*(1), 24-42.

Mihret, D., James, K., & Mula, J. (2010). Antecedents and organizational performance implications of internal audit effectiveness. *Pacific Accounting Review, 22*(5), 224-252. doi:https://doi.org/10.1108/01140581011091684

Morrill, J., Morrill, C., & Kopp, L. (2012). Internal Control Assessment and Interference Effects. *Behavioral Research in Accounting, 24*(1), 73-90.

Mueller, F., & Carter, C. (2007). We are all managers now: Managerialism and professional engineering in UK electricity utilities. *Accounting, Organizations and Society, 32*(1-2), 181-195.

Nerantzidis, M. (2015). Measuring the quality of the "comply or explain" approach. *Managerial Auditing Journal, 30*(4/5), 373-412. doi:http://dx.doi.org/10.1108/MAJ-08-2014-1060

Nerantzidis, M. (2016). A multi-methodology on building a corporate governance index from the perspectives of academics and practitioners for firms in Greece. *Corporate Governance, 16*(2), 295-329. doi:https://doi.org/10.1108/CG-08-2015-0107

Nerantzidis, M., Filos, J., Tsamis, A., & Agoraki, M.-E. (2015). The impact of the Combined Code in Greek soft law: Evidence from 'comply or explain' disclosures. *International Journal of Law and Management, 57*(5), 445-460. doi:https://doi.org/10.1108/IJLMA-05-2014-0036

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering, 48*, 500-506. doi:https://doi.org/10.1016/j.proeng.2012.09.545

Oussii, A. A., & Taktak, N. B. (2018). The impact of internal audit function characteristics on internal control quality. *Managerial Auditing Journal, 33*(5), 450-469. doi:https://doi.org/10.1108/MAJ-06-2017-1579

Pizzini, M., Lin, S., Vargus, M., & Ziegenfuss, D. (2015). The impact of internal audit function quality and contribution on audit delays. *Auditing: A Journal of Practice & Theory, 34*(1), 25-58.

Prawitt, D., Smith, J., & Wood, D. (2009). Internal audit quality and earnings management. *The Accounting Review, 84*(4), 1255-1280.

Pugliese, A., Bezemer, P. J., Zattoni, A., Huse, M., Van den Bosch, F. A., & Volberda, H. W. (2009). Boards of directors' contribution to strategy: A literature review and research agenda. *Corporate Governance: An International Review, 17*(3), 292-306.

Ramamoorti, S. (2003). *Internal Auditing: History, Evolution, and Prospects.* IIA, Institute of Internal Auditors. Retrieved from https://na.theiia.org/iiarf/Public%20Documents/Chapter%201%20Internal%20Auditing%20History%20Evolution%20and%20Prospects.pdf

Rezaee, Z., Sharbatoghlie, A., Elam, R., & McMickle, P. L. (2002). Continuous auditing: Building automated auditing capability. *Auditing: A Journal of Practice & Theory, 21*(1), 147-163.

Rud, O. P. (2001). *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management.* New York, USA: John Wiley & Sons Inc.

Salehi, M., & Bahrami, M. (2017). The effect of internal control on earnings quality in Iran. *International Journal of Law and Management, 59*(4), 534-546.

Sarens, G., & Abdolmohammadi, M. (2011). Factors associated with convergence of internal auditing practices: emerging vs developed countries. *Journal of Accounting in Emerging Economies, 1*(2), 104-122.

Sarens, G., De Beelde, I., & Everaert, P. (2009). Internal audit: a comfort provider to the audit committee. *The British Accounting Review, 41*(2), 90-106.

Schneider, K., & Becker, L. (2011). Using the COSO model of internal control as a framework for ethics initiatives in business schools. *Journal of Academic and Business Ethics, 4*, 1-18.

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Data Warehousing*, *5*, 13-22.

Sheikh, R. E. (2011). *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications.* Business Science Reference.

Sirikulvadhana, S. (2002). *Data Mining As A Financial Auditing Tool - M.Sc. Thesis in Accounting.* Swedish School of Economics and Business Administration.

Soh, D., & Martinov-Bennie, N. (2011). The internal audit function, perceptions of internal audit roles, effectiveness and evaluation. *Managerial Auditing Journal, 26*(7), 605-622. doi:https://doi.org/10.1108/02686901111151332

Steinberg, R. M., Everson, M. E., Martens, F. J., & Nottingham, L. E. (2004). *Enterprise Risk Management - Integrated Framework: Executive Summary.* Committee of Sponsoring Organizations of the Treadway Commission.

Sultana, N., Singh, H., & Van der Zahn, J.-L. (2015). Audit committee characteristics and audit report lag. *International Journal of Auditing*, *19*(2), 72-87.

Turnbull, G. R. (2005). *REVISED GUIDANCE FOR DIRECTORS ON THE COMBINED CODE.* London: FINANCIAL REPORTING COUNCIL - INTERNAL CONTROL. Retrieved from https://www.frc.org.uk/getattachment/fe1ba51a-578d-4467-a00c-f287825aced9/Revised-Turnbull-Guidance-October-2005.pdf

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Elsevier.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., . . . Steinberg, D. (2008). Top 10 algorithms in data mining. *International Conference on Data Mining - ICDM. 14*, pp. 1-37. Springer-Verlag London Limited. doi:10.1007/s10115-007-0114-2

Zain, M. M., Subramaniam, N., & Stewart, J. (2006). Internal Auditors' Assessment of their Contribution to Financial Statement Audits: The Relation with Audit Committee and Internal Audit Function Characteristics. *International Journal of Auditing*, *10*(1), 1-18. doi:https://doi.org/10.1111/j.1099-1123.2006.00306.x

Zhu, Q., & Lin, X. (2007). Top-Down and Bottom-Up Strategies for Incremental Maintenance of Frequent Patterns. *Emerging Technologies in Knowledge Discovery and Data Mining*. doi:https://doi.org/10.1007/978-3-540-77018-3_44