Multi-criteria segmentation of syndicates trading in the Lloyd's market

Mémoire d'actuariat

Muriel Blum
November 2017

Mémoire à tendance professionnelle
Collège des Ingénieurs Paris
PartnerRe Paris

Academic Supervisor: Olivier Lopez
Supervisor at PartnerRe: Matthew Hazzard

Abstract

Key words: Lloyd's market, reinsurance, syndicate, capital provider, segmentation, linear model building

The Lloyd's market in London is a unique marketplace which enables insurance and reinsurance syndicates to access worldwide and specialist risks thanks to a global license and distribution network, its brand name as well as stable financial strength ratings.

The team "Capital at Lloyd's" in the international reinsurance group PartnerRe is active as a capital provider for syndicates trading in the Lloyd's market. In order to identify the right opportunities for such long-term commitments while optimizing the return on capital, it is desirable to segment the syndicates according to their expected profitability in the future as well as other characteristics.

By compiling and studying data from different sources as well as reviewing existing analyses of syndicates, this project aims at refining the understanding of success factors and establishing a method for systematically analyzing and comparing all the syndicates trading in the Lloyd's market.

Professional intuition is combined with statistical methods to identify a set of characteristics of syndicates that deserve a closer look when searching for the most profitable syndicates.

Two new segmentation methods are presented as alternatives to syndicate segmentation approaches based on two existing external scoring systems. One has been developed based on the outcomes of linear model building presented in this work. The other one is much more based on professional intuition and on strategic preferences of PartnerRe.

Hypothetical portfolios composed of syndicates segmented with these four different methods as well as a random selection are compared in terms of performance. Two approaches to portfolio construction and three different measures for profitability are compared.

The methods for syndicate segmentation that were developed in this project are not clearly better than existing methods. Nevertheless, the development of these methods represents an independent confirmation of the rankings established by external sources. When looking at those rankings in the future, their predictions can be used with an increased confidence.

The Lloyd's market is an attractive marketplace for a capital provider who is able to identify and seize suitable opportunities through a complex process which starts from a multi-criteria segmentation of syndicates, results in successful capital deals and requires the expertise of experienced professionals all along the way.

Résumé

Mots clés : Marché des Lloyd's, réassurance, syndicat, fournisseur de capital, segmentation, modélisation linéaire

Le *Lloyd's market* à Londres est une place de marché unique au monde qui permet aux entreprises d'assurance et de réassurance d'accéder à des risques globaux et spécialisés grâce à un réseau de licences et de distribution global, à sa réputation et à la stabilité de sa solidité financière.

L'équipe *"Capital at Lloyd's"* au sein du groupe de réassurance PartnerRe est active en tant que fournisseur de capital pour des syndicats opérant dans le *Lloyd's market*. Dans le but d'identifier les meilleures opportunités pour ce type d'engagements à long terme tout en optimisant le retour sur investissement, il est important de segmenter les syndicats en fonction de leur profitabilité attendue pour le futur ainsi qu'à d'autres caractéristiques.

En compilant des données de différentes sources et des analyses existantes de syndicats, ce projet a pour objectif d'affiner la compréhension des facteurs de succès et d'établir une méthode pour systématiquement analyser et comparer tous les syndicats actifs du *Lloyd's market*. L'intuition professionnelle est combinée avec des méthodes statistiques pour identifier un ensemble de caractéristiques des syndicats qui devraient être prises en compte lors de la recherche des syndicats les plus profitables.

Deux nouvelles méthodes de segmentation des syndicats sont présentées ici comme alternatives aux approaches de segmentation basées sur deux systèmes de notation extérieurs. L'une d'entre elles a été développée en utilisant les résultats de modèles linéaires présentés dans ce travail. L'autre méthode est basée sur des réflexes professionnels et sur les préférences stratégiques de PartnerRe.

Des portefeuilles hypothétiques composés de syndicats segmentés avec ces quatre différentes méthodes ainsi qu'avec une sélection aléatoire sont comparées les unes aux autres en termes de performance. Deux différentes manières de construire un portefeuille et trois différentes mesures de profitabilité sont comparées. Les méthodes de segmentation des syndicats qui ont été developpées dans le cadre de ce projet ne sont pas clairement meilleures que les méthodes existantes. Néanmoins, le développement de ces méthodes représente une confirmation indépendante des systèmes de notation établis par des sources externes. Ce sera donc avec une confiance renforcée que les predictions de ces derniers pourront être prises en compte dans le futur.

Le *Lloyd's market* est une place de marché intéressante pour un fournisseur de capital s'il est capable d'identifier et de saisir les opportunités adéquates dans un processus commençant par la segmentation multi-critère des syndicats, demandant l'expertise de professionnels expérimentés tout au long du processus et résultant de prises de capital réussies.

Contents

Acknowledgments

My special thanks go to Matthew Hazzard and Michel Büker for giving me the opportunity to work on this *Mémoire*. They have supervised and guided me during the work on this topic as well as all along the mission that I realized in their team in the framework of the MBA program at Collège des Ingénieurs.

Matthew has always been patient in answering my questions and explaining me the particularities of the Lloyd's market and he has supported my work by sharing his expertise.

I also thank Olivier Lopez for all his efforts as the academic supervisor of this *Mémoire*. I appreciated the scientific guidance that he gave me during our meetings.

Special thanks also go to Sili Lu at PartnerRe who has helped me with the software R and with related questions. Furthermore, I am grateful to Karl Whitehead, Adrien Haegel and many other colleagues at PartnerRe who have supported me directly or indirectly during my time at PartnerRe.

I would also like to thank Bruno Boulay and Hélène Feau who have accompanied my mission and my work on the *Mémoire* on the side of the Collège des Ingénieurs.

Last but not least, I thank my family and friends for their continued support.

PartnerRe

Notes about Notation

The decimal separator in this document is the point. The comma is used to indicate thousands.

The following list spells out the abbreviations used in this document and indicates the corresponding page numbers.

Abbreviations

AIC Akaike Information Criterion. 18, 38, 46, 47, 72

APCL Argenta Private Capital Limited. 15, 16, 20, 29, 50–52, 54, 60, 64, 70

BIC Bayes Information Criterion. 18, 38, 46, 72

CAL Calendar year. 8, 23–26, 28, 29, 50, 51, 65, 69, 77

CM Corporate Member. 1, 4, 5, 32, 62

ECA Economic Capital Assessment. 10, 12

FAL Funds At Lloyd's. 9–13, 62

GPW Gross Premiums Written, gross of reinsurance. 5, 7, 21, 32

LCR Lloyd's Capital Return. 10, 12

MA Managing Agent. 1, 4, 5, 7, 10, 12, 14, 15, 23, 26, 30, 62

NEP Net Earned Premiums, net of reinsurance. 21, 22, 32, 34

NPW Net Premiums Written, net of reinsurance. 21, 32

RDS Realistic Disaster Scenario. 4, 15, 22, 60, 62

RITC Reinsurance To Close. 4, 7, 8, 16

ROC Return On Capital. 2, 9, 14, 24, 28, 51, 54, 55

SBF Syndicate Business Forecast. 9, 16

SCO Syndicate Continuity Opinion. 14, 15, 54

SCR Solvency Capital Requirement. 10, 12, 15

SPA Special Purpose Arrangement. 5, 16, 25, 42

SRL Syndicate Research Limited. 14, 15, 19–21, 29, 32, 50–52, 54, 60, 64, 65

TPC Third Party Capital. 1, 5, 8, 16, 25–27, 29, 32, 51, 53, 62, 63, 69

YOA Year Of Account. 7, 8, 15, 23–26, 28, 29, 51, 52, 69, 77

1 Introduction

The time-honored Lloyd's market in London is a unique marketplace which enables insurance and reinsurance companies to access worldwide and specialist risks thanks to a global license and distribution network, its brand name as well as stable financial strength ratings.

The Lloyd's market has a capital framework which cannot readily be duplicated elsewhere combined with an overarching, consistent performance management framework across all key aspects of a business. Lloyd's "Chain of Security" and its capital model are interesting to look at from an actuarial point of view.

For PartnerRe as an international reinsurance group, it is attractive to participate in this market. In particular, PartnerRe's Paris-based team responsible for "Capital at Lloyd's" (formerly known as "Lloyd's Net Quota Share and Multiline") is active as a capital provider for syndicates trading in the Lloyd's market.

Syndicates are created on an annual basis to write insurance and reinsurance business. They are backed financially by so-called Members and operated by Managing Agents (MAs). The capital providers, i.e. Members, can be wealthy individuals (called Names) or Corporate Members (CMs). The role of a MA is to run a syndicate on behalf of the capital providers. This basic structure of the Lloyd's market is schematically illustrated in Figure 1 (see page 6).

While a large part of the CMs are owned by (re)insurance companies which typically also own MAs, there is also Third Party Capital (TPC) embedded into such vehicles. All the capital provided by Names (through a variety of vehicles) should also be considered as TPC because Names are third parties. However, in this document, given the marginal interest in the role played by Names, the term TPC is usually used as a shorthand for *corporate* third party capital, meaning capital invested by companies through CMs.

Those companies which own and control a MA in support of their own syndicate are providing so-called "corporate capital" (first party, not third party) through their dedicated CM. If all of the capital backing a syndicate is corporate capital, the syndicate is called "fully aligned". On the other hand, if there is a panel of capital providers of a syndicate, including either Names and/or (corporate) TPC, the syndicate is "non-aligned" [1].

PartnerRe is currently providing TPC to several different syndicates and

1. These definitions of alignment are very specific and should not be confused with the colloquial use of the word "alignment" for example in the sentence "the interests of the two parties are well aligned".

is aiming at further increasing the number of such investments over the next few years.

Therefore, in order to optimize the Return On Capital (ROC), it is desirable to identify the syndicates which are most likely to be profitable in the future. Even though a syndicate is an annual venture, the participation in a syndicate is oftentimes seen as a long-term commitment, making it even more important to select the right opportunities.

There are a multitude of factors that play a role in the selection of investment opportunities, including many "soft" factors relating to people in the market. However, it would be desirable to complement and corroborate the experience of the investment managers in PartnerRe's "Capital at Lloyd's" team with a systematic quantitative analysis.

By integrating data from different sources as well as existing analyses of syndicates, this project aims at refining the understanding of success factors and establishing a method for systematically analyzing and comparing all the syndicates trading in the Lloyd's market.

Professional intuition is combined with statistical methods to identify a set of characteristics of syndicates that deserve a closer look when searching for the most profitable syndicates. Practical considerations such as data availability, methodological artifacts and simplicity of the method are also discussed.

Subsequently, by defining selection criteria and corresponding weightings, segmentation methods for everyday use can be obtained. One will be based on the outcomes of a linear model building process while another one will be based on professional reflexes and strategic preferences of PartnerRe.

A comparison of the added value of these segmentation methods compared to alternative approaches as well as a critical discussion thereof complete the practical part of this *Mémoire*.

Given the specificities of the problem and the particularities of the Lloyd's market, the topics relating to this *Mémoire* are not well covered in body of literature formed by past *Mémoires*. The references in this text will therefore mainly point to documentation surrounding the Lloyd's market. We believe that the contents are nevertheless relevant for the French *Institut des Actuaires*.

2 Theoretical Part

The theoretical part presents some relevant concepts about the Lloyd's market before looking at two existing approaches to syndicate scoring and finally briefly recalling some theory about linear modeling.

2.1 Particularities of the Lloyd's market

In view of the peculiar structure of the Lloyd's of London market, this theoretical part begins by giving some background, introducing some key concepts and clarifying some terminology.

2.1.1 History

More than 300 years ago, a coffee house in London, the *Edward Lloyd's Coffee House* emerged as a hub for information about shipping. The latest news about marine adventures were highly relevant to the first individuals providing shipping insurance - a very profitable but risky business. Therefore, during the first half of the 18th century, the Lloyd's establishment became a hotspot of marine underwriting and its influence reached a global scale.

In the 1760s, the underwriting diversified to some non-marine lines of business, in 1769, a restructuring resulted in the "New Lloyd's" and in 1773 a man called John Julius Angerstein originated the concept of a lead underwriter, meaning that others followed him by underwriting the same policies at the same rates.

In 1811, the creation of a network of Lloyd's agents further increased the flow of information to Lloyd's and thereby consolidated its reputation and expertise. By the middle of the 19th century, financial security was reinforced by requiring Members to put up a deposit to support their underwriting. By 1870, the concept of large syndicates was introduced, thereby increasing the size of the lines that could be written and standing up to the growing competition from companies outside of Lloyd's.

The first Lloyd's reinsurance policy on American risks was written in the 1880s. In the 1890s, the role of brokers became increasingly important.

At the start of the 20th century, a financial audit process was introduced at Lloyd's as a consequence of some scandalous financial failures. Moreover, the pricing of policies became much more risk-based thanks to a wide range of data and studies about hurricanes, earthquakes, etc. Also at the beginning of the 20th century, business began to be written through delegated authorities in other countries.

The San Francisco earthquake in 1906 resulted in several other novelties, including excess of loss reinsurance. In the 1920s, a case of an enormous debt of an underwriter initiated the creation of the "Central Fund", adding even more financial security to policyholders. The globalization emerging in the 1960s made Lloyd's business truly international.

In 1986 the New Lloyd's building at One Lime Street was opened. A series of problems in the 1980s and 1990s relating to some very risky covers, unlimited liability and other issues resulted in "the most turbulent and traumatic time" [1] in Lloyd's history. The subsequent restructuring and the establishment of "Equitas", a vehicle into which all pre-1993 business was transferred by Reinsurance To Close (RITC), marked a new beginning for a more robust and modern Lloyd's.

In 1994, the first CMs began underwriting. In 2001, Lloyd's became subject to oversight by the new Financial Services Authority. In the 2000s, several other new concepts and entities were introduced, including the Franchise Board (responsible for underwriting and risk management standards across the market), annual accounting (supposed to replace three-year accounting, see Section 2.1.3) and Realistic Disaster Scenarios (RDSs) (requiring syndicates to model their expected losses for certain major disasters).

The current development of Lloyd's is guided by the Vision 2025 that was launched in 2012.

This brief summary of the history of Lloyd's was based on [1], where many more details can be found.

2.1.2 Structure

Lloyd's is not a company but a marketplace for brokers and syndicates. Insurance and reinsurance business is written when the right connections are made between underwriters and brokers representing policyholders.

Lloyd's describes itself as *the world's leading market for specialist insurance*.

The typical flow of business, illustrated in Figures 1 and 2, begins with the policyholder who requires some insurance cover.

On behalf of the policyholder, a broker will then place the risk in the market. The brokers are the intermediaries between the insured and the underwriters of the syndicates. With their specialist knowledge of the market, the brokers approach the underwriters, introduce the business and receive 15-30% of premium. There are about 200 broker firms in the Lloyd's market and they bring business from about 200 countries and territories.

Furthermore, there are so-called coverholders who place risks. They are also known as managing general agents (not to be confused with MAs). Under the terms of a binding authority issued by a MA, these companies can enter into insurance contracts in the name of the Members of the corresponding syndicate. Some of the Lloyd's brokers can act as coverholders.

The MAs manage the syndicates, i.e. are responsible for all operational decisions and processes. The people actually running the syndicates are thus people working for the corresponding MAs - a syndicate as such is not really a company or any other legal entity. Nevertheless, the syndicates are said to write insurance risks.

The risks are actually borne by the Members who provide the capital to back the syndicates. As capital providers, these Members are exposed to the liabilities resulting from underwriting, at least up to their limit of liability. The Members are also those who benefit from profits made by the syndicates. The MAs are of course taking their share of the profits as well - their remuneration is based on a combination of fees and profit commissions.

The two basic types of Members are Names and Corporate Members (CMs), corresponding to individuals and to companies respectively. Among the CMs, there are those which are dedicated to a specific syndicate and those which participate in several syndicates. TPC CMs are typically in the second category, i.e. they back different syndicates. The portfolio of such CMs and the corresponding interactions with MAs are managed by Members' Agents. For Names, it is also essential to have recourse to the services of a Members' Agent in order to be active in the market.

Finally, there is the Corporation of Lloyd's which supports the market in various ways. It is for example responsible for the overarching performance management framework, for market services and for the interaction with regulatory authorities. Additionally, the Corporation of Lloyd's is managing the central assets of Lloyd's which form the third link in the Lloyd's "Chain of Security", see Section 2.1.5.

The number of syndicates and MAs has been evolving over time. Table 1 gives an idea of the number of these entities in the past few years. In 2017, there are 96 active syndicates, 57 Managing Agents and 4 Members' Agents (3 main ones).

Some syndicates are so-called Special Purpose Arrangements (SPAs) which depend on another syndicate and its MA. As they are taking a part of the book of business of their parent syndicate, some accounting adjustments have to be made when determining capacity, premium or other figures at the Lloyd's market level.

The Lloyd's market as a whole has a capacity of about 30 billion GBP (see Table 1), with "capacity" referring to an upper limit of premiums gross of reinsurance but net of acquisition costs (see also Section 2.1.4).

In contrast, the Gross Premiums Written, gross of reinsurance (GPW) as it appears in usual financial statements is gross of reinsurance and gross of acquisition costs. As the acquisition costs and the amount of unutilized

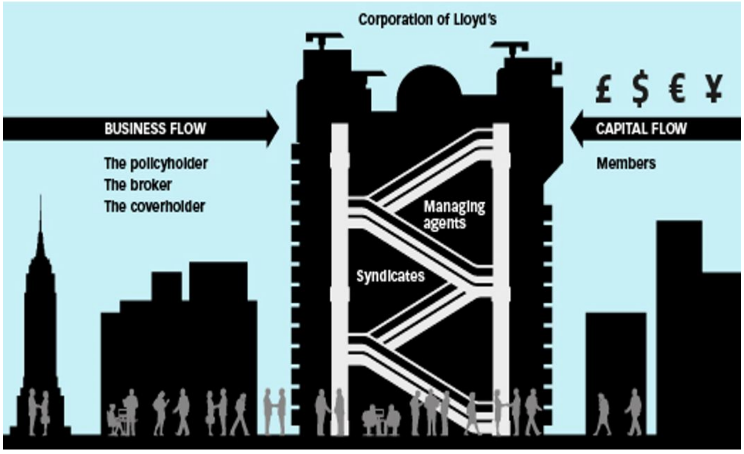


Figure 1 – Schematic representation of the Lloyd's market showing the flows of business and of capital as well as the Corporation of Lloyd's which oversees the market. Source [13].

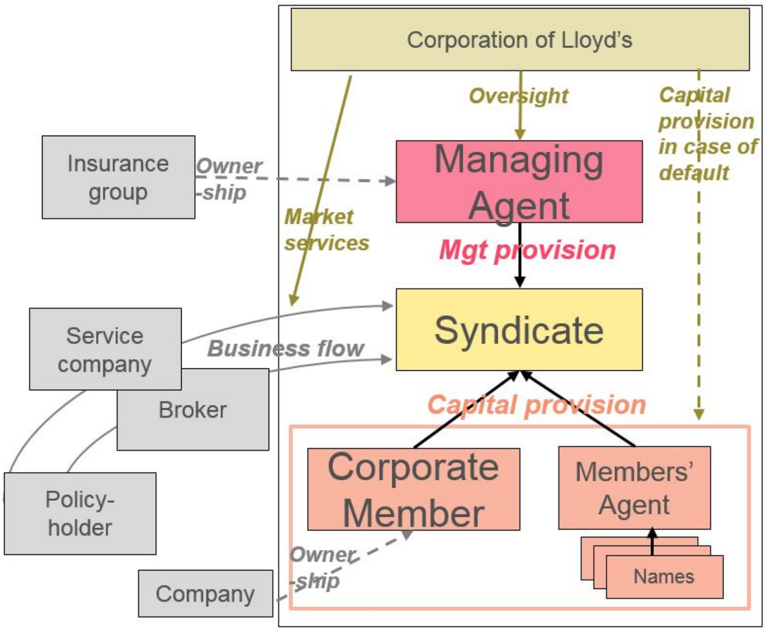


Figure 2 – Schematic representation of the structure of the Lloyd's market showing the different internal and external stakeholders and their interactions. For simplicity, only one syndicate is depicted.

capacity are roughly compensating each other, the GPW is oftentimes quite close to capacity, but the two should not be confused. The total Lloyd's GPW was for example GBPm 30,538 in 2016 and GBPm 27,545 in 2015 (see Lloyd's pro forma financial statements).

Table 1 – Number of syndicates, MAs and Members' agents active in the Lloyd's market over the past years. Source [9].

	2017	2016	2015	2014	2013
Lloyd's Capacity (GBPm)	30,198	27,609	26,266	26,527	24,998
Syndicates	96	98	99	95	90
Managing Agents	57	59	59	56	55
Members' Agents	4	4	4	4	4

	2012	2011	2010	2009	2008
Lloyd's Capacity (GBPm)	24,167	23,314	22,951	18,136	16,106
Syndicates	90	92	87	85	80
Managing Agents	56	56	54	53	51
Members' Agents	4	4	4	4	4

2.1.3 Accounting

Originally, the syndicates at Lloyd's have always used three year accounting. A Year Of Account (YOA), also called underwriting year, includes every risk *written* during the corresponding calendar year with all the premiums and claims related to it. For business written in long-tail lines of business such as casualty, the claims will appear quite far in the future, sometimes several years later than the risk was written. Even for short-tail lines of business such as property, it is clear that for policies underwritten towards the end of a given year, the premiums and claims will materialize over a period longer than the underwriting year.

The concept of three year accounting thus corresponds quite well to the period at risk of business underwritten during a given year and with annual policies. Historically, the three year accounting goes back to the maritime logic that Lloyd's would pay claims if a ship did not return after three years.

After the end of the third year, the syndicates work on the RITC. This is a contract between the Members of the closing YOA and the following YOA. The latter take on all prior liabilities in exchange for a premium that they receive from the members of the closing YOA. The YOA can thereafter be called a "closed" year.

RITC thus corresponds to passing on the reserves for liabilities successively from one YOA to another. The terms of an RITC have to be fair and equitable according to Lloyd's rules. This means that whenever there are too many uncertainties in the RITC, the corresponding YOA will be left "open" until estimates are more reliable.

Nowadays, Lloyd's is using annual accounting. This makes it more comparable to other (re)insurance players. The basis for annual accounting is the Calendar year (CAL).

However, many syndicates still use three year accounting, especially the syndicates with TPC because in their case, the participation of such capital providers (which can participate with different shares on different YOAs) requires that different underwriting years can be separated from each other.

As Figure 3 illustrates, the CAL and the YOA are not the same.

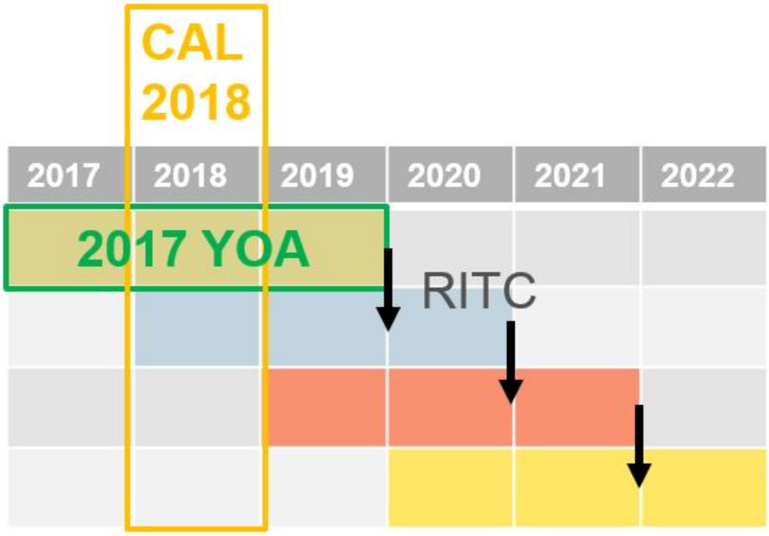


Figure 3 – Schema illustrating the difference between YOA and CAL and the closing of each YOA into the subsequent YOA by the RITC mechanism.

2.1.4 Capacity vs. Capital

Besides accounting periods, another potential source of confusion in the Lloyd's market it the concept of capacity.

The so-called stamp capacity (or "capacity" or "stamp") of a syndicate is an upper limit of premium income net of acquisition costs (but gross of reinsurance).

Acquisition costs consist of commissions paid to brokers and other related expenses.

To prevent any confusion between capital and capacity, we usually spell out capacity rather than abbreviating it.

The planned premium income that a syndicate declares in its business plan (known as Syndicate Business Forecast (SBF)) which has to be approved by Lloyd's is usually lower than the capacity. This is to allow room for exchange rate movements or increasing premium in the aftermath of a sudden change in market conditions. Syndicates can only underwrite more premium than their capacity if they get permission from Lloyd's.

The capital that Members put up to support the underwriting of the syndicate(s) in which they participate is known as Funds At Lloyd's (FAL).

The amount of FAL depends on the riskiness of the business, but it is in most cases lower than the capacity that it is backing.

The FAL is lodged at Lloyd's in the form of cash, shares or bank guarantees. In some special cases, the funds can be held at syndicate level and are known as FIS.

These assets represent the second link in Lloyd's "Chain of Security", see Section 2.1.5.

When speaking about the *profitability* of a syndicate, it is common to consider the ratio of the financial result to the capacity of the syndicate.

$$\text{Profitability} = \frac{\text{Profit in GBPm}}{\text{Capacity in GBPm}}$$

This allows for a straightforward calculation and for inter-syndicate comparability of profitability, whereas the ratio of profit to capital would depend on the Member under consideration.

In particular, those Members who invest in a portfolio of syndicates may enjoy a diversification benefit that lowers the amount of capital that they need to put up compared to the sum of equivalent participations by different Members.

The ratio between capital and capacity is thus individual to each Member. More details about capital setting will be given in Section 2.1.6.

When it comes to making general assumptions, here we use a ratio of:

$$\frac{\text{FAL}}{\text{Capacity}} = 60\%$$

The Members' Agent Hampden assumes a ratio that changes over time: 40% for 2001-2007, 45% for 2008-2011 and 50% from 2012 onwards [16].

Our 60% assumption is thus rather conservative.

The ROC of a participation in Lloyd's syndicates is thus roughly a factor $1/60\% = 1.67$ bigger than the profitability expressed as profit per capacity.

2.1.5 Chain of Security

Lloyd's capital structure is quite unique and powerful. It provides very high financial security to the policyholders all while being quite capital-efficient for Members.

There are three links in the so-called "Chain of Security", as schematically illustrated in Figure 4:

1. Syndicate level assets
2. Members' Funds at Lloyd's (FAL)
3. Central assets

The first link in the Chain of Security are the premiums collected by the syndicates which are held in trust at the syndicate level. They are the first resource for settling valid claims of policyholders.

Second, there is the capital held in trust funds at Lloyd's for each Member, i.e. the FAL. In cases where syndicate assets are insufficient to meet liabilities, the MA makes a call on Members' FAL.

The third link in the Chain of Security consists of different types of central assets of Lloyd's. The "Central Fund" is available at the discretion of the Council of Lloyd's to meet any Member's insurance liabilities. Similarly, there are also the corporation assets and the subordinated debt which are available as central assets. Moreover, Lloyd's retains the right to call a contribution of 3% of capacity from all syndicates, meaning that there are another GBP 900m providing further financial security to policyholders.

2.1.6 Capital model

The calculation to determine the required amount of FAL is done at the level of each Member. Members participating in different syndicates may get a diversification benefit from their portfolio of participations.

The underlying "Lloyd's Internal Model" is not public. Members can use the "Member Modeller" to calculate their capital requirements based on their portfolio. The different steps of the capital setting process - which is known as Economic Capital Assessment (ECA) - are schematically illustrated in Figure 5 and briefly described below. More information about the ECA can be found in Lloyd's ECA guidance manual [3].

Among the inputs for the Lloyd's Internal Model, there are the numbers about the Solvency Capital Requirement (SCR) of each of the trading syndicates. The process by which Lloyd's collects information about each syndicate's capital position is the so-called Lloyd's Capital Return (LCR).

Each syndicate is obliged to determine its SCR in two different manners: on a one-year basis and "to ultimate". The SCR on a one-year basis (SRC1)

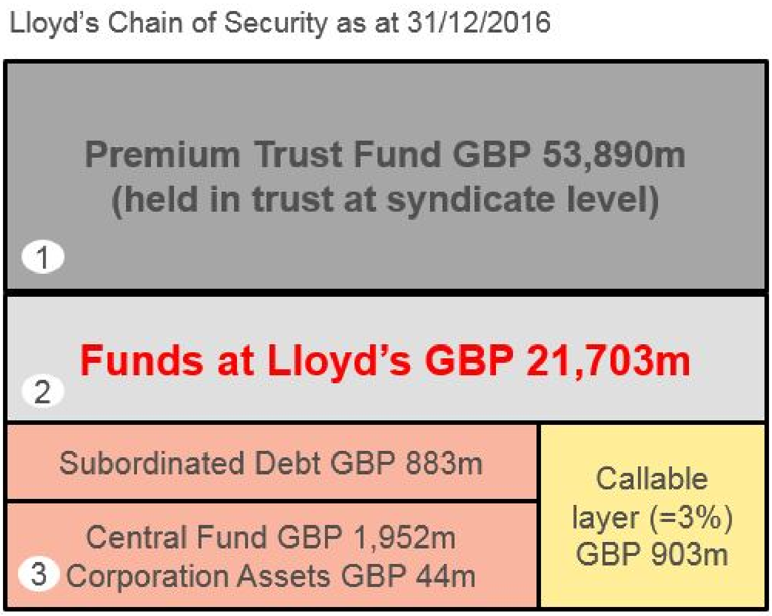


Figure 4 – Schematic representation of Lloyd's unique capital structure, called the "Chain of Security". There are three links, namely the syndicate level assets, the Members' Funds at Lloyd's (FAL) and the central assets. Adapted from [2].

is consistent with Solvency II regulations and its calculation is a regulatory requirement. However, Lloyd's considers the ultimate SCR (uSCR) as the more appropriate risk measure [4]. The uSCR takes into account the risks over the full lifetime of the liabilities as assessed at a confidence level of 1:200.

Along with the uSCR and SCR1 numbers, each syndicate, i.e. its MA, submits a "SCR Documentation" which allows Lloyd's to understand the operational details of the syndicate's internal model and its compliance with regulation.

Next, the uSCR is uplifted by a certain percentage to yield the ECA. Currently the uplift is 35% (confirmed for 2018), but this number might change in the future as it is subject to annual review by Lloyd's Franchise Board. The SCR together with the Lloyd's uplift and any potential Solvency II accounting adjustments is called the **ECA of the syndicate**.

The minimum regulatory **capital requirements for a Member** are calculated from the ECA values of the different syndicates that this Member is backing. Adjustments due to diversification or concentration of risks within the portfolio are made. This is the stage where the Lloyd's Internal Model comes into play. The ECA values of the syndicates are combined to yield the Members' ECAs.

Members' FAL requirements are calculated twice a year. This process of capital setting is called "coming into line".

A Member's FAL must be at least 40% of its total capacity. This is now required for all Members while in the past there were exceptions for members writing mainly Motor business. The FAL that the Member actually puts up is thus not necessarily equal to the ECA number coming from the model.

Moreover, Lloyd's will introduce a limit on Tier 2 capital (e.g., letters of credit and bank guarantees) in Members' FAL. The limit will be at 90% of ECA for the 2019 YOA and then further decreasing (80% for 2020 and 75% for 2021) [5]. Solvency deficits will have to be fully covered by Tier 1 capital (e.g., cash, bonds and equities).

Data collected through the LCR are a direct input into the Lloyd's Internal Model and are also used to calibrate it. There is a series of updates of the "Member Modeller" software during autumn corresponding to the different stages at which Lloyd's receives and approves data and documents provided by the MAs. The parameters of the Lloyd's Internal Model are also reviewed and updated based on historical data.

Furthermore, there are always some new focus areas in which the capital setting process should be improved. Currently, attention is directed for example to the lack of adjustment of the internal models to widening of terms and conditions. The additional risk of wider terms and conditions is currently not

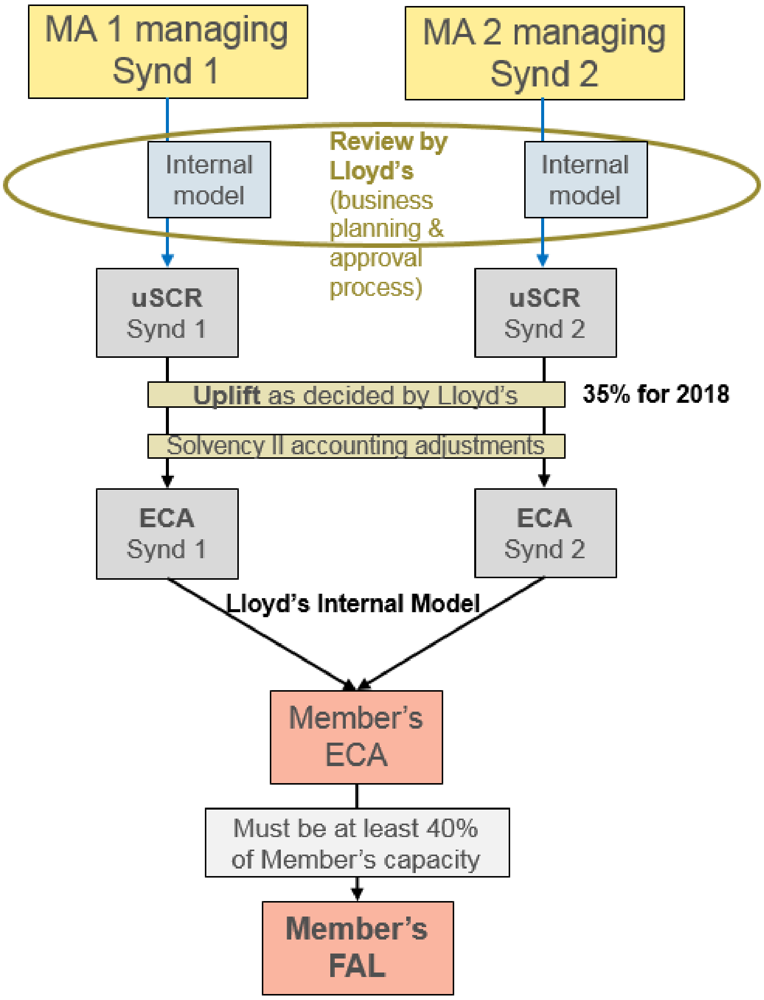


Figure 5 – Schematic representation of the capital assessment processes which at the end indicate how much FAL each Member has to put up.

reflected in increased capital requirements and thus there is no internal cost to such underwriting decisions. Another example is the topic of Cyber lines of business where the parametrization and validation processes are under review and the MAs are required to submit additional information.

2.2 Existing analyses of syndicates

Before starting to develop our own framework for scoring the performance of syndicates, we summarize the approaches taken by other organizations.

2.2.1 Syndicate Research Limited

Syndicate Research Limited (SRL) is a company whose main activity is to provide research about the syndicates. In its own words, it is a company "committed to providing independent research on all active syndicates trading at Lloyd's" and it "offers 100% coverage of individual Lloyd's syndicates as well as the Lloyd's market as a whole" [6].

The team at SRL has many years of cumulated Lloyd's expertise.

SRL publishes a variety of data and analyses about the syndicates trading in the Lloyd's Market. Their profiles about each syndicate and each MA provide a set of relevant data and descriptions at one glance.

As a synthetic evaluation, SRL assigns a so-called Syndicate Continuity Opinion (SCO) to a syndicate.

The SCO is based on a "Scorecard" which is based on quantitative factors complemented by qualitative considerations. The resulting "Aggregate Score" is then transformed into the SCO by making further adjustments due to qualitative information if necessary.

Specifically, the information taken into account in the Scorecard are the following (weighting indicated in parentheses):

- 9 year average ROC (70%)
- 2 year average ROC (5%)
- 2 year average Combined Ratio (5%)
- Absolute number of distinct material business lines (15%)
- 2 year average of Expenses (incl. forex) % Net Premiums Earned (5%)

Each of these factors is evaluated on the C-, C, C+, B-, B, B+, A-, A, A+ scale, corresponding to numerical values 1 to 9.

After calculation of the weighted average (with the weights as indicated above), the Scorecard indicator is subject to further adjustment based on the "Other Considerations".

These Other Considerations are based on the following factors and result in an adjustment of the score within the range indicated in parentheses:

— Franchise value (± 0.5)
— Underwriting cycle management (-0.5)
— Percentage of new business lines (-0.5 or -0.25)
— Management stability (-0.5)
— Group/External support (from -1 to $+1$)

Finally, the resulting score is mapped back to the C- to A+ scale, yielding the SCO.

A Scorecard is available for 75 syndicates (representing 94% of the market's capacity) and a SCO is given for 45 syndicates (74% of the market's capacity).

2.2.2 Members' Agents

Argenta Private Capital Limited (APCL) is one of the three main Members' Agents in the Lloyd's market. In order to facilitate the investment decisions of the Members utilizing their services, they publish a brochure with syndicate profiles in which a variety of ratings are given to each syndicate with which they collaborate.

It is important to notice that the data and analyses published by APCL do not cover all the syndicates trading in the Lloyd's market, but only the syndicates open to APCL Members (28 for the 2017 YOA, 27 for the 2018 YOA). This is a big difference to SRL and will be further discussed in Section 3.2 about data availability.

The overall rating that APCL gives to a syndicate ranges from D through C, C+, B, B+, A to A+.

They do not disclose the full details of their methodology (and they write that "the rating includes a degree of subjectivity" [12]), but the main factors taken into account in their different scorings are nevertheless indicated.

The "Risk Rating" is designed to indicate the likelihood of a large loss for the syndicate relative to the market as a whole. It is based on a combination of the SCR, the RDS, the volatility of past results, the exposure to reinsurance failure and the quality of the MA. The Risk Rating is given on a verbal scale (Lower, Medium, Higher, Very High).

Several other ratings are given on a scale from 0 to 10. Table 2 shows what kind of information the different scores are based on.

Comments about Auction Price:

Some of the capacity is held on a secure basis for the corresponding Members, i.e. it cannot be taken away from them by the MA of the respective syndicate. The origin of this so-called "freehold" capacity was an agreement

Table 2 – Factors for the different APCL ratings and the corresponding methodology.

APCL Rating	Factors	Calculation
Capital	Marginal capital requirement	For adding a line of 25k to a portfolio of 1m GBP
Catastrophe	1 in 30yrs Aggregate Exceedance Probability figures (latest SBF)	(details undisclosed)
Tail	a) Claims paid at 3yrs % ultimate claims b) RITC premium % net premium	Some combination of both (details undisclosed)
Cost	Adjusted avg auction price (previous year)	10 if not traded at auctions (e.g. SPAs)
Scarcity	a) Capacity available at auctions b) Capacity provided by TPC	Ratio of the two

that Members negotiated with Lloyd's in the early 1990s after the "Reconstruction and Renewal" project. Since 1995, this capacity owned by Members can be traded via auctions.

The auction price can give an indication of how successful a syndicate is.

However, as it is only the freehold capacity that is traded on auction, this measure is available only for a limited number of syndicates (e.g. only 19 syndicates in 2016).

2.3 Theory about linear modeling

In a linear model, the dependence of a dependent variable y on multiple independent variables $x_1, ..., x_p$ is described as

$$y_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) + \varepsilon_i$$

for each observation i.

The independent error terms ε_i are assumed to follow a normal distribution with mean 0 and equal variance σ^2.

Ordinary Least Square regression is a statistical method to determine the parameters of the linear model for a given dataset. If the predictor variables $x_1, ..., x_p$ are given, the calculation of the least square estimates of the coefficients β, written as $\hat{\beta}$, is straightforward and the resulting relationship can be used to predict y from $x_1, ..., x_p$.

$$E[Y] = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_p X_p$$

The sum of the squares of the errors (residuals) is written as SS_{Resid}. Similarly, SS_{Regr} stands for the sum of the squares of the regression and SS_{Total} for the total sum of squares. With $\bar{y}$ designating the mean value of the observed y_i values, we can write

$$SS_{Resid} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad SS_{Regr} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad SS_{Total} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

However, the fundamental problem in this work is that it is not known in advance which independent variables should be included in the model. Therefore, the selection of predictors to be included in the model is the main part of the analysis.

Generally, we can distinguish between two ways of selecting variables [21]:

1. Comparing **all possible** subsets of the pool of candidate predictors with a certain criterion for quality of fit.
2. Using a search algorithm and a criterion for quality of fit for finding an optimal model in an **iterative** way.

The first approach can only be implemented if the number of potential predictors p is low because the number of subsets scales as 2^p (because each predictor can either be in a subset or not (2 possibilities) with all of these p decisions being independent). Hence, the second approach will be taken here (for 19 independent variables, $2^{19} = 524'288$).

There are different search algorithms for this iterative approach, namely
— Forward selection
— Backward elimination
— Stepwise regression

A basic way to describe the quality of fit of a model is the R-squared value R^2, also called the coefficient of determination:

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{SS_{Regr}}{SS_{Total}} = 1 - \frac{SS_{Resid}}{SS_{Total}}$$

However, the R^2 will continue to improve when more variables are added. That is why it is more useful to look at the *adjusted* R^2 value, written as R_a^2, which takes into account the number of variables [23]:

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2)$$

with n designating the number of observations and p the number of predictors. The number of parameters of the model is $p+1$ due to the intercept.

As each criterion has its advantages and its limitations, there are many criteria that have been developed for comparing the quality of fit of linear models. Two other criteria that will be used in this work are the following:
- Akaike Information Criterion (AIC)
- Bayes Information Criterion (BIC) which is also known as Schwarz' Bayesian Criteria (SBC)

These are criteria that consist of a combination of two elements: on the one hand the maximum likelihood estimates of the model parameters (on a logarithmic scale) and on the other hand an adjustment penalizing for the number of predictors.

The lower the value, the better. According to [18], the expressions for these criteria in a regression setting are:

$$\mathrm{AIC} = \text{constant} + n\log(SS_{Resid}) + 2p$$

$$\mathrm{BIC} = \text{constant} + n\log(SS_{Resid}) + \log(n)p$$

with n designating the number of observations and p the number of predictors as above [2].

BIC penalizes a high number of variables more strongly than AIC, meaning that its use will result in a final model with fewer retained variables.

2. It should be noted that the last term of the expression is given differently depending on the source, with sometimes $p+1$ or $(p+1)+1$ instead of p. It is the number of β_i *parameters* of the model that is $(p+1)$, i.e. the *predictors* plus the intercept (in some sources written as "p" unlike here). The actual formulas are $\mathrm{AIC} = n + n\log 2 + n\log(\frac{SS_{Resid}}{n}) + 2(q+1)$ and $\mathrm{BIC} = n + n\log 2 + n\log(\frac{SS_{Resid}}{n}) + \log(n)(q+1)$ for a model with $q = p+1$ parameters (β_0 to β_p) and σ^2 taken into account as well. However, the difference of AIC or BIC values between two models - the quantity of interest in an iterative search algorithm - will be the same regardless of these subtleties. More details can be found in [20].

3 Practical Part (Application)

3.1 Objectives

This work aims at providing some insights into the success factors of syndicates and at translating those into a segmentation method that can assist the investment decisions of PartnerRe as capital provider to syndicates. This implies that the following main steps should be taken:

1. Make an inventory of available data and find a way to use them in a practical way (Section 3.2).
2. Search for factors that are associated with syndicate profitability (Sections 3.3 and 3.4).
3. Combine the identified factors to a scoring system that can be used to rank syndicates (Section 3.5).
4. Compare the performance of this scoring system to other available methods in terms of benefits and costs (Section 3.6).

The first objective is closely related to other projects of PartnerRe's "Capital at Lloyd's" team which go beyond the scope of pure syndicate segmentation and are therefore not discussed here.

For achieving the second objective, a combination of expert knowhow and mathematical modeling was considered as most promising. Linear model building is thus used to complement the intuition of professionals with many years of experience.

The methodologies and discussions relating to the third objective are mostly confidential and thus presented in Appendix II. Nevertheless, a brief high level summary of this part is given in the main document.

The conclusions from the fourth objective are in the end placed in the practical context of PartnerRe as a capital provider to Lloyd's syndicates.

3.2 Availability and quality of data

This section concerns the objective number 1 set above (Section 3.1), namely making an inventory of available data in order to find ways to use them in a practical way.

Quantitative and qualitative data are discussed separately.

3.2.1 Quantitative data

Data about Lloyd's syndicates was combined from several different sources:
— SRL Syndicate Peer Data, an Excel file published quarterly
— Lloyd's Statistics, a set of Excel files published annually

— APCL brochure, a series of syndicate profiles published annually
— SRL Syndicate Profiles, a series of syndicate profiles updated at irregular intervals

In general, all these sources provide high quality data.

APCL data is not really usable for a segmentation method that should cover all the syndicates because it is only available for the syndicates with which APCL is dealing (fewer than 30 syndicates).

The main weakness of the data is the lack of data for certain years and for certain syndicates.

For the data that comes from financial statements, it is supposed that the accuracy of the data is high.

Accuracy might be less high for variables that are based on some undisclosed analysis by the data providers. For example the variable "Lloyd's Business as % Group Total" is based on some analysis by SRL which we did not try to reproduce and double-check.

Selected quantitative data used for linear model building:

As preparation for the linear model building process described in Section 3.4 below, a subset of the available data was chosen. For linear model building, it is desirable to have a high number of data points. That is why a certain set of variables was chosen for which data availability is high and/or which are considered as relevant in modeling profitability. These selected elementary variables are presented in Table 3.

The annual accounting profit and loss statement of each syndicate is available in the Lloyd's Statistics. This is therefore a good starting point for a series of variables. Attention has to be paid to the signs of the different variables. Table 3 therefore contains a column that describes which variable is given with which sign.

3.2.2 Qualitative data

There are various types of qualitative data that one might want to take into account as well.

Subjective impressions of PartnerRe underwriters who have practical experience in dealing with syndicates have been collected as -/0/+ along with explanatory comments. These data are not used for the model building or the segmentation, but they can be displayed next to the ranked syndicates as additional information.

In the different descriptive text paragraphs in the syndicate profiles provided by SRL and APCL, there is also a lot of useful information. However, practically speaking, the extraction of this information is not straightforward.

Table 3 – Selected variables for linear model building. Labels with S refer to SRL data while labels with L refer to the Statistics. The labels are introduced to facilitate the definition of candidate predictors based on variable transformations, see Table 4. Net Earned Premiums, net of reinsurance (NEP) is different from Net Premiums Written, net of reinsurance (NPW), which can be calculated by adding L1 and L2.

Label	Description	Unit	Typical sign
S1	Capacity	GBPm	always >0
S2	Lloyd's Business as % Group Total	%	always >0
S3	Aligned Dedicated Share of Syndicate Capacity %	%	always >0
L1	Gross Premiums Written (GPW)	GBPm	>0 income
L2	Outward reinsurance premiums	GBPm	<0 expense
L3	Net Earned Premiums (NEP)	GBPm	>0 income
L4	Syndicate investment return	GBPm	>0 income
L5	Claims paid gross amount	GBPm	<0 expense
L6	Claims paid reinsurers' share	GBPm	>0 income
L7	Change in provision for claims gross amount	GBPm	<0 expense
L8	Change in provision for claims reinsurers' share	GBPm	>0 income
L9	Operating expenses acquisition cost	GBPm	<0 expense
L10	Operating expenses administrative expenses	GBPm	<0 expense
L11	Profit / (loss) on exchange	GBPm	>0 or <0
L12	Profit Pre-Tax	GBPm	>0 or <0

Different approaches based on keyword searches or other methods are imaginable and could be reconsidered in future work.

3.3 Preliminary considerations and analyses

In order to achieve the objective number 2 set above (Section 3.1), namely the identification of factors associated with syndicate profitability, we will first start by discussing some concepts and pitfalls relating to profitability. This will pave the way for making reflexions around linear model building as described in Section 3.4 below.

3.3.1 What exactly are we looking for?

The main objective for a capital provider is to identify syndicates which have a high profitability all while having a low volatility.

A high profitability is typically observed in syndicates that write a lot of catastrophe business, but it is undesirable to have too much of a concentration on such books because they are also more volatile.

The method of choice here to avoid high volatility syndicates is to apply a filter after having ranked the syndicates by expected profitability. In principle, it would be imaginable to build two separate segmentation methods with one aiming at maximizing profitability and the other aiming at minimizing volatility. One could then take the intersection of the two rankings, i.e. the syndicates which are ranked high in both of them.

However, this approach was discarded because the presence of a low number of high-volatility syndicates can be accepted if the investment portfolio is quite large. Given than PartnerRe is now participating in more than 10 syndicates, it can accept small shares of syndicates with a high volatility.

That is why the focus was laid purely on profitability in this segmentation work. The considerations relating to volatility remain to be applied at a later stage, with different measures of volatility being potentially involved. The standard deviation of profitability over a given time horizon (for example 4 or 8 years) is the most obvious measure of volatility. An alternative approach is to look at the RDS percentages that the syndicates provide in their business plans, which is in practice hindered by limited access to data. It can also be a combination of the two that is used to create a score for filtering out high-volatility syndicates if necessary.

We would therefore like to identify measurable characteristics of a syndicate which are associated with high profitability.

Many of the characteristics of a syndicate are directly or indirectly reflected in certain variables for which data is available. The following list is not exhaustive but gives a few examples of factors which might affect the profitability of a syndicate:

— The size of the syndicate resulting in economies of scale => consider the variable capacity or a variation thereof.
— The experience accumulated within the syndicate => consider the age of the syndicate.
— The riskiness of the business written, because as expected, a high volatility is usually associated with high profitability => consider the RDS percentages.
— The efficiency of the operations of the syndicate => consider the expense ratio (administrative and acquisition expenses divided by NEP).

3.3.2 Different ways of defining profitability

Before going further, it has to be clear what we mean by profitability.

It has already been defined that profitability is the financial result divided by the capacity of a syndicate, as opposed to the capital (see Section 2.1.4 above).

Next, it has to be clarified which accounting system we are considering. As described in Section 2.1.3 above, there are some particularities of accounting at

Lloyd's, namely the difference between CAL and YOA. Both types of data are used in this work depending on the context, so it is important to distinguish them.

One can also look at different time horizons over which profitability can be averaged in order to get a meaningful value despite the cyclic nature of the (re)insurance market.

These degrees of freedom taken together result in a multitude of potential definitions of profitability of a syndicate. They will be further explored in the "Comments about taking averages" below.

To add to the complexity, it is not only the profitability of closed accounting periods that could be considered as relevant, but also the forecasted profitability of currently open YOAs. The most recent developments of a syndicate could actually be of high importance for the segmentation.

Comments about forecasts of syndicates' results:

Every quarter starting from the end of the fifteenth month of a YOA, i.e. as of quarter 5, the MAs are establishing forecasts of their syndicates' results. Hence, for any analysis based on YOA results, the question arises whether or not to take into account the forecasted results of the open YOAs. If they are not taken into account, the latest developments are ignored. If they are taken into account, the data availability between different syndicates diverges because only the non-aligned syndicates are obliged to make these forecasts public.

Also, it should be kept in mind that these are forecasts and that by nature it is difficult to forecasts the returns of (re)insurance - a major loss event might result in a significant deterioration of forecasted results. This is also why forecasts are not even published before the fifteenth month of a YOA.

Moreover, the first forecasts come with a wide range between worst case and best case. The midpoint of these two estimates is usually different from a syndicates point-estimate. However, the latter, which is the more interesting one, is not public information.

Finally, the trends observed in the development of forecasts over time often follow the underwriter's maxim "good years get better and bad years get worse" [14]

All of these aspects should be kept in mind when thinking about profitability.

Comments about taking averages:

When looking at profitability - expressed as percentage of capacity unless indicated differently - it seems useful to look at long-term averages. However, the question arises of how to choose the most suitable period.

Given the cyclic nature of the (re)insurance business, averages should be taken over an extended period of time which is covering different phases of the cycle.

Here, averages of profitability are usually taken over 8 years. In some cases, averages are taken over 4 years. This relates to data availability and/or other considerations. In Section 3.6.1 where different segmentation methods will be compared, the rationale behind 4-YOA averages for that purpose is explained.

It should be noted that simple averages are taken even though the capacity of some syndicates is changing from one year to another.

Another noteworthy point is the fact that the average is actually based on fewer years than it suggests if a syndicate has been trading only for part of the considered time horizon (for example a start-up syndicate) or if data availability is partial.

In addition to 8-year and 4-year averages, long-term averages are considered in some cases in order to take into account as much data as available (namely CAL data that is available back to 2004).

In that case, two methods of taking a long-term average are applied:

1. Taking a simple average over the entire period.
2. Calculating rolling averages over 5-year periods and then taking a simple average of these averages.

The second method gives a reduced weight to the beginning and the end of the period, i.e. the distant and the most recent past respectively.

The second method is chosen to facilitate evaluation of PartnerRe's Group objectives fixed for ROC on any 3- to 5-year period.

Obviously, the question arises how closely these different averages are associated with each other. Figure 17 in Appendix I shows three plots that show the relation between the average of rolling 5-CAL averages over the period 2004-2016 and

— the simple 13-CAL (2004-2016) average
— the 8-YOA (2009-2016) average and
— the 4-YOA (2013-2016) average

respectively. Every syndicate for which data is available appears as a data point. As expected, the correlation is the highest in the first case and the lowest in the third.

Yet another pitfall with averages concerns those averages that are taken across time and across syndicates. For example, one might want to compare the average profitability over the last 8 years of aligned syndicates to the one of non-aligned syndicates.

However, due to the pattern of data availability in combination with the trends in results, it matters how we take such averages.

Namely, as illustrated in Figure 6, we observe that:

- Results are generally better for earlier years, following the general market trend (symbolized by ▲, Year3 being the earliest year in the Figure).
- Results are generally worse for young syndicates, for example the latest SPAs (▼).

Consequently, we can observe that the value we obtain when first averaging over the years (separately for each syndicate) and then averaging this result over the syndicates is generally lower than the value obtained vice versa, i.e. $avg_{synd}(avg_{year}) < avg_{year}(avg_{synd})$.

This should be kept in mind when considering such averages.

Figure 6 – Illustration of how the pattern of data availability in combination with the trends in results creates a bias in the averages, meaning that the order in which averages across syndicates and across year are taken matters. Year1 stands for the latest year, Synd1 for the oldest syndicate.

3.3.3 Comparing groups of syndicates

Besides the continuous variables that are introduced in Section 3.2 and used as a basis for linear model building in Section 3.4, there are also some categorical variables that could be associated with profitability.

It would be useful to know whether the syndicates belonging to a certain group are more profitable than the others and if yes to explore potential reasons. For instance, several of the Members' Agents claim that the syndicates with which they deal are outperforming the market. More generally, it is commonly affirmed that the syndicates which have TPC, either from Names or from corporate TPC providers, are achieving a higher profitability than the fully aligned syndicates (see for example [11]).

Here, we have looked at the syndicates which have a least some *corporate* TPC compared to those which do not.

For these two groups, the two charts in Figure 7 show the average profitability for each of the YOAs and CALs respectively.

Overall, the graphs and the average profitability of each group across the years suggest that the TPC syndicates[3] are indeed outperforming the others. It is however necessary to perform an actual statistical analysis because the variances are not given on the graphs and because, as we have seen in Section 3.3.2, there are some pitfalls with taking averages both across syndicates and across years.

One-way ANOVAs[4] performed for separate CALs revealed that the difference between the two groups is significant in certain years but not in others. For the CAL 2013, the TPC syndicates have a significantly different (better) performance than the others at the 0.05 confidence level ($p = 0.029$) while for the CAL 2016, the difference is far from being significant ($p = 0.915$).

When combining the data from all the different CALs in the period 2004-2016, the difference between the TPC syndicates (206 data points) and the others (618 data points) is significant ($p = 0.015$).

The trend is less clear in the YOA data which includes forecasted data and where data availability is lower because some of the aligned syndicates do not publish YOA figures. When combining the data from the different YOAs in the period 2009-2016 (including forecasts), a p-value of 0.059 is found when testing the hypothesis of an equal average profitability of the TPC syndicates (147 data points) and the others (247 data points). However, if the forecasted 2015 and 2016 YOAs are excluded, the TPC syndicates are very clearly outperforming the others ($p = 0.002$, 101 and 225 data points respectively).

Overall, we can conclude that the presence of corporate TPC has a positive influence on profitability. Therefore, we will use variables such as "TPC" and "Alignment" when building linear models, see Table 4 on page 31.

Another example of a group that was considered is the group of syndicates managed by so-called turn-key Managing Agents. These are MAs that run a syndicate for a third party in exchange for a fee for their services. The underlying idea was that syndicates managed by a turn-key MA could be more performant than the others thanks to the pooling of expertise over several different syndicates.

However, their average profitability is found to be neither better nor worse than the average profitability of all other syndicates (ANOVA with combined CAL data 2004-2016, $p = 0.439$). Hence, we refrain from searching for some continuous variable related to this syndicate characteristic that could be included as a predictor variable for profitability.

3. As before, we do not always explicitly write that we are looking at *corporate* TPC (ignoring Names).

4. For two groups, one-way ANOVA is equivalent to the t-test for difference in means.

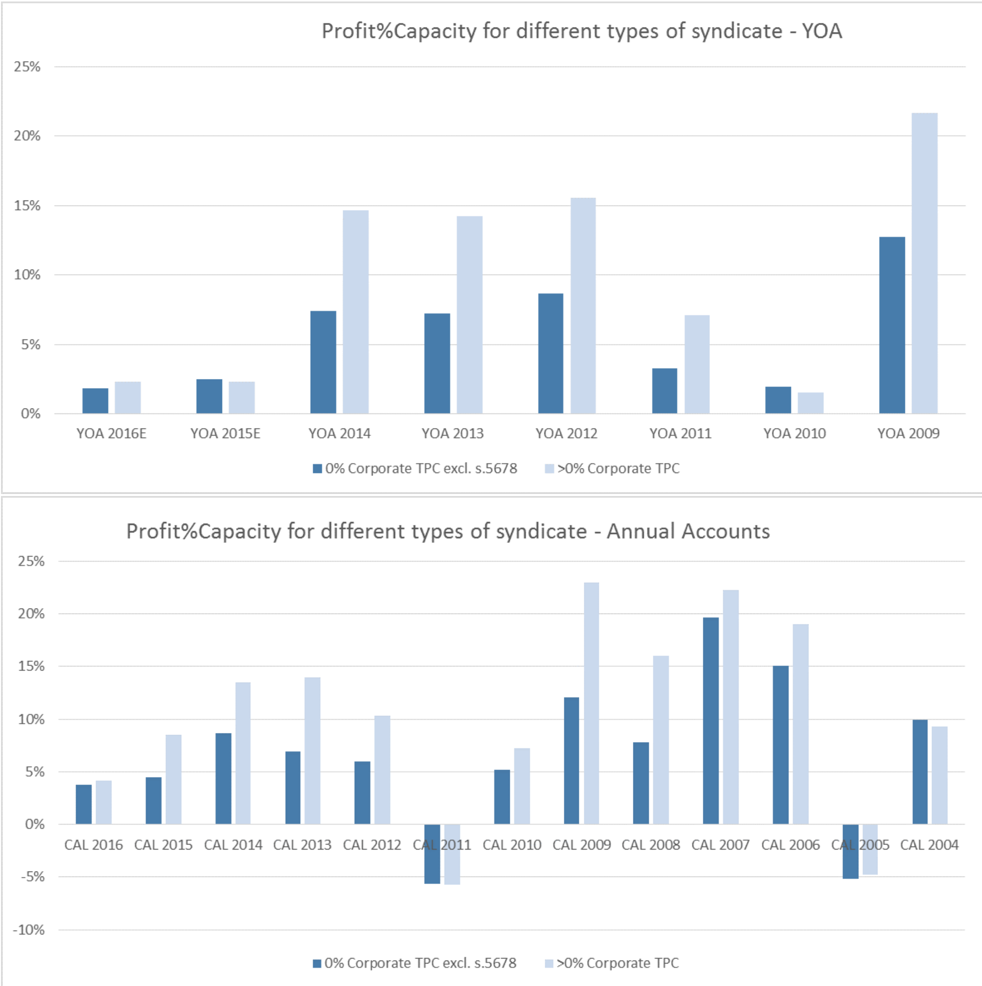


Figure 7 – Comparing the performance of syndicates with and without corporate TPC. The impression conveyed by the graph - that the TPC syndicates are outperforming the others - is confirmed using ANOVA. One syndicate had to be excluded from the data due to an artifact.

3.4 Building linear models to understand drivers of syndicate profitability

Following the preliminary considerations above, we can now dive into the linear model building process aimed at identifying factors associated with syndicate profitability (objective number 2 set in Section 3.1 above).

The response variable that should be modeled is profitability, i.e. the financial result of a syndicate as a percentage of capacity, because this is what one would like to maximize as a capital provider [5].

The motivation for building linear models [6] for profitability is two-fold:

— On the one hand, the linear models and the process of obtaining them should shed some light on the interdependence of different variables and thereby contribute to a refined understanding of syndicate dynamics.

— On the other hand, the linear models should provide a basis for selecting relevant variables to be considered in a segmentation method.

This section begins by describing the approaches to linear model building that were taken in this work, including considerations about the choice of dataset. Second, the results of model building with the original data as training set are presented. Third, the imputation of data and the corresponding refined results are exposed. Finally, a synthesis of the insights gained through the different models is presented and ideas for future research are described.

3.4.1 Model building approaches

The process of linear model building applied here consists of the following general steps:

1. Prepare the dataset which serves as training set
   — Select candidate predictors
   — Impute data where necessary
   — Look at correlation matrix to get a first idea of associations
2. Run an algorithm for linear model building
   — Select type of algorithm
   — Select criterion for comparison of models
3. Interpret the resulting set of retained predictors
4. Compare the outcomes of models built with different algorithms and trained with different datasets

The software R [17] was used to build series of nested linear models and compare their quality of fit using different criteria. In order to be used in R, the datasets were exported as csv files from the Excel files, making sure that the precision of the data was not lost.

Due to the difference between YOA and CAL and the data presented correspondingly, there were two main possible datasets that were considered as a basis for linear models:

5. To be precise, it is the ROC that we would like to maximize, but as explained above, data is available for capacity rather than for capital.

6. We write about linear models in the plural form because depending on the selected criterion and the dataset, different models can be found as optimal solutions.

1. 8-year averages of YOA data
2. CAL data for several years separately

Some preliminary attempts to building linear models were made with the averaged YOA data. The reason why YOA data is not used separately by years is that it includes forecasted data. It therefore seems more meaningful to work with averaged YOA data to circumvent two main issues:

— Forecasts are only published by non-aligned syndicates, meaning that for aligned syndicates no data points would be available for the latest YOAs.
— By nature, forecasts might be significantly different from reality, which is why such data points were not used in their own right.

Overall, CAL data was finally considered as much more useful because it allows for more data points. Hence, the selection of variables presented in Table 3 above (page 21) was used as a basis for systematic linear model building.

These elementary variables were combined and transformed in order to obtain variables that are suspected to be associated with profitability in a linear way. This is a process for which expert knowhow is crucial. Based on discussions with the experienced professionals in PartnerRe's "Capital at Lloyd's" team, a selection of candidate predictors was made. Table 4 presents the chosen candidate predictors and the formulas for obtaining them from the variables for which data is directly available.

Some of the variables are quite obviously connected to profitability given that we have the following relationship [7]:

$$\begin{aligned}\frac{\text{Profitability} \cdot \text{Capacity}}{\text{NEP}} =& 1 - \text{NLR} - \text{Acquisition} - \text{Admin} \\ & + \text{InvestmentReturn} + \text{Forex}\end{aligned}$$

Other variables are chosen because they are suspected or known to be associated with profitability, for example reinsurance efficiency or TPC (see Section 3.3.3 above).

Looking at the SRL and APCL methodologies in Section 2.2.1, we see some other factors that might be interesting to take into account. Some information related to the lines of business written and their long- or short-tail

7. This equation corresponds to the structure of the profit and loss statement L12=L3+L4+L5+L6+L7+L8+L9+L10+L11, see Table 3, which is then normalized by NEP and takes account of the sign changes as defined through the formulas in Table 4.

orientation is captured in the LTOS. As for the measures related to capital, it is unfortunately the data availability issue that prevents us from using similar measures.

Even though all kinds of syndicate characteristics could potentially be associated with syndicate profitability, we are only interested in those that would somehow make sense from the point of view of an experienced professional. As a counter-example, let's imagine that we were looking at the number of letters in the name of the MA of the syndicate. In a hypothetical scenario, we might find a positive correlation between the length of the name and the profitability of syndicates. It could then be discovered that the name is on average longer for larger syndicates, for example because they tend to have undergone some mergers and thus concatenations of names in the past. Despite the association of profitability and name length, the more meaningful variable in that case would be the capacity of the syndicate.

Table 4 – Response variable (Profitability) and candidate predictors for linear modeling along with the formulas for obtaining them from the variables for which original data is directly available (see Table 3) and with comments about the sign of these predictors. The labels of the variables are used to refer to them in the text.

Label	Meaning	Way to calculate it	Sign
Profitability	Profit % Capacity	$\frac{L12}{S1}$	>0 for profit or <0 for loss
BasisMarket	Average Lloyd's market return on capital	from other source	>0 for profit or <0 for loss
Capacity	Capacity	$S1$	always >0
LogCapacity	Logarithm of Capacity (variable transformation)	$\log(S1)$	>0 if Capacity above 10 GBPm
GroupSupport	Lloyd's Business as % of Group Total	$S2$	>0 percentage
Alignment	Aligned Dedicated Share of Syndicate Capacity in %	$S3$	>0 percentage
TPC	Corporate Third Party Capital in % (share of capacity)	from analysis	>0 percentage
Age	Number of years of existence (counter starts at 1993)	from other source	always >0
LTOS	Long-Tail Orientation Score (0 to 1)	from analysis	always >0
InvestmentReturn	Investment return % NEP	$\frac{L4}{L3}$	>0 for profit or <0 for loss
ReinsEff1	Reinsurance efficiency [recoveries/spend]	$\frac{L6+L8}{-L2}$	>0 ratio
ReinsEff2	Reinsurance efficiency [(recoveries-spend)/GPW]	$\frac{L6+L8-(-L2)}{L1}$	>0 for profit or <0 for loss
Acquisition	Acquisition costs % NEP	$\frac{-L9}{L3}$	>0 expense ratio
Admin	Administrative expenses % NEP	$\frac{-L10}{L3}$	>0 expense ratio
Forex	Foreign exchange rate gain or loss % NEP	$\frac{L11}{L3}$	>0 gain ratio or <0 loss ratio
PremiumRetention	Premium retention [NPW/GPW]	$\frac{L1-(-L2)}{L1}$	>0 percentage
WrittenVsEarned	(NPW-NEP)/NEP	$\frac{L1-(-L2)-L3}{L3}$	>0 for growth or <0 for shrinking
NLR	NLR=net claims incurred % NEP	$\frac{-(L5+L6+L7+L8)}{L3}$	>0 ratio
GLR	GLR=gross claims incurred % gross earned premium	$\frac{-(L5+L7)}{L1\frac{L3}{L1-(-L2)}}$	>0 ratio
PercentagePaid	Gross claims paid % gross claims incurred	$\frac{L5}{L5+L7}$	>0 ratio

For those candidate predictors in Table 4 which do not result from a straightforward calculation based on variables for which data is available, the following paragraphs provide insights into the data sources and methodologies.

The variable BasisMarket was included into the dataset even though it is not a characteristic of a syndicate. The data for BasisMarket was taken from the Lloyd's Statistics where pro-forma financial statements for the market as a whole are presented. The available data is a measure of profit per capital rather than profit per capacity, but this is of minor relevance here because the purpose of this predictor is anyways just to give a baseline for the evolution of the market. The idea is that the underlying trends in the market could potentially explain a part of the variability of the syndicate results that could not be accounted for by other factors. As we will see later, this is not really the case, but it was worth an attempt.

The percentage of corporate TPC was obtained by analyzing data from SRL about the top 5 capital providers of a syndicate. By excluding all the dedicated CMs as well as all the Members' Agents and their vehicles, the share of corporate third parties was determined. The data does unfortunately not allow to capture small shares of corporate third parties which are either below the top 5 or hidden within the participation of a Members' Agent. Nevertheless, this analysis gives an idea of the percentage of capacity of a syndicate that is backed by corporate TPC.

The age of the syndicates was calculated based on another data file provided by SRL.

The calculation of the LTOS is based on the split of business by line of business (data from SRL) in combination with a classification of each of these lines of business as 0 (short-tail business), 0.25, 0.5, 0.75 or 1 (long-tail business). For example, the line of business "Goods in Transit" has the coefficient 0 while "Accident & Health" is assigned a coefficient 1. This analysis certainly has its limitations, but it can serve as a rough indication of the long-tail orientation of a syndicate.

The calculation of the GLR should also be explained because it involves the assumption that the ratio between earned and written premium is the same for gross as for net. Hence, the ratio of NEP to NPW is used to scale the GPW in the denominator. The numerator is the sum of the paid claims and the reserved claims, both on a gross basis, i.e. the gross claims incurred.

Among the 19 candidate predictors, there are some pairs of predictors that are obviously correlated with each other. More insights into correlations between the potential predictors will be given below in the sections where Pearson correlation matrices are discussed.

Including highly correlated predictors in linear model building causes redundancies. However, for each case there was a rationale for including such a pair of variables among the candidate predictors.

- Capacity was complemented by LogCapacity because it seemed reasonable to assume that the size of a syndicate would influence profitability in a logarithmic rather than linear way.
- NLR and GLR are two ways of looking at a syndicate's underwriting result. As the teams within a syndicate responsible for underwriting and for buying reinsurance/retrocession are usually separate, both loss ratios could be relevant. The GLR is more indicative of the pure underwriting performance. The NLR is also taking into account the performance of the reinsurance program. As a capital provider, the NLR is what finally matters (along with the other components that form the final result). Nevertheless, as we will see later, a combination of GLR and a measure of reinsurance efficiency will be more predictive for profitability than NLR.
- Two different ways of looking at reinsurance efficiency were used. While ReinsEff1 is only looking at the percentage of recovered expenses, the ReinsEff2 is putting the absolute gain or loss through the reinsurance program in relation to the total premium volume. A syndicate that buys only a small amount of reinsurance for a very specific book of business and happens to bee lucky (i.e. does gain rather than lose money on its reinsurance program) can score outstandingly well on ReinsEff1 but cannot reach extreme values on ReinsEff2.

Therefore, despite the suspected redundancies, these candidate predictors were included into the model building process because it was not clear in advance which one of the two in each pair is more meaningful.

The results of the model building processes showed that there was usually only one or the other of such a pair of predictors appearing in the retained model. Therefore it was acceptable to include redundancies in the beginning.

What about variable transformations in the process of linear model building?

The candidate predictors as defined in Table 4 are already expressions that combine elementary variables based on industry know-how. For example, the different expenses are normalized by NEP and candidate predictors such as PremiumRetention or PercentagePaid are quotients of other elementary variables.

This reduced the need for variable transformations during linear model building.

As described above, the Capacity was transformed to LogCapacity. For other variables however no transformation seemed justified. Quadratic and higher-order polynomial dependencies would not be expected in the context of profitability. Even though they might increase the quality of fit of the model, uncovering such dependencies would not be particularly useful for the research question at hand.

Additionally, using variable transformations that would increase the fit of the model because they would replicate the intrinsic connection of variables according to the equation on page 29, for example $\frac{\text{NLR}\cdot\text{Capacity}}{\text{NEP}}$, are uninteresting (NEP is not even among the candidate predictors).

3.4.2 Models trained with original data

This section discusses the first series of models based on the unimputed dataset with 19 candidate predictors. Even tough the dataset has 620 data points, only 121 of them are complete data points and thus usable for linear model building [8] .

It is important to notice that within the original dataset before imputation, the complete data points all appear within the three latest years (2014-2016) because the variables GroupSupport and Alignment are not available for earlier years. Despite this major limitation, the linear models built with this original, unimputed data are briefly discussed here in order to later contrast them with those obtained for the imputed dataset covering 8 consecutive years.

As we will see, the imputed dataset will give rise to much more conclusive linear models. A cross-market-cycle view and a correspondingly large number of datapoints thus seem to be important when analyzing syndicate profitability.

We could also expect that the variable BasisMarket shows up in the optimal models based on 8 years but not on 3 years, which is true to some extent (see Figure 12).

The correlation matrix (Pearson correlation) of all variables (1 response variable and 19 predictor variables) was investigated. It is presented in Table 5 and pronounced correlations are highlighted.

Notable observations include the following:

— As expected mathematically, LogCapacity and Capacity are correlated.

— Alignment and TPC are negatively correlated, which is expected because the more aligned a syndicate is, the less capacity it can allocate to corporate TPC providers.

— The high correlation of GLR and NLR among each other (positively) and with Profitability (negatively) is not surprising either because the loss ratios are the main component in the profit and loss statement of a (re)insurance company.

8. With the exception of a particular model building approach which will be labeled e) and is described on page 39.

— The correlation of PremiumRetention and ReinsEff2 is related to the definition of ReinsEff2 which makes it, in contrast to ReinsEff1, sensitive to the percentage of ceded premium. Syndicates are usually losing money on the reinsurance they buy (but gaining a reduced volatility in return), so the ReinsEff2 is usually negative. However, it is of lower magnitude, i.e. closer to being positive, when the syndicate ceded only a small share of its business, i.e. has a high PremiumRetention.
— At the first glance more surprising is the positive correlation between WrittenVsEarned and Admin. It indicates that for example a fast-growing syndicate - for which the written premium exceeds the premium earnings - has rather high administrative expenses. The main component of administrative expenses usually being salaries of staff, this seems to suggest that more staff is hired before a growth phase. On the opposite, the reduction of staff expenses would be associated with a reduction of written premium preceding a reduction of earned premium.
— Finally, the negative correlation between GLR and PercentagePaid is also quite surprising. The lower the GLR, the more of the claims are paid rather than reserved for. One could think that this relates to short-tail business which could on average be more slightly more profitable than long-tail business (see also the -0.08 between LTOS and Profitability as well as the 0.18 and 0.15 between LTOS and GLR and NLR respectively). However, it is important to note that the LTOS is not correlated to the PercentagePaid (0.02). Another potential explanation is that a high GLR appears in those cases where there are some large claims which are, unlike attritional losses, not directly paid for but reserved for while being disputed.

Table 5 – Based on the data before imputation ($n = 121$ complete data points): Pearson correlation matrix for the dependent variable Profitability and the 19 independent variables. Coefficients above 0.5 are highlighted in red, those below -0.5 in blue.
Prof=Profitability, BM=BasisMarket, Cap=Capacity, LogCap=LogCapacity, Group=GroupSupport, Align=Alignment, Invmt=InvestmentReturn, RE1(2)=ReinsEfficiency1(2), Acq=Acquisition, PrRet=PremiumRetention, WVsE=WrittenVsEarned, PPaid=PercentagePaid.

	Prof	BM	Cap	LogC	Group	Align	TPC	Age	LTOS	Invmt	RE1	RE2	Acq	Admin	Forex	PrRet	WVsE	NLR	GLR	PPaid
Prof	1.00	0.15	0.28	0.25	0.13	0.08	-0.31	0.46	-0.08	0.39	-0.14	-0.15	-0.39	-0.32	0.15	0.03	-0.34	-0.84	-0.81	0.42
BM	0.15	1.00	0.06	0.04	-0.04	0.07	-0.04	-0.07	-0.02	0.14	-0.07	-0.03	-0.02	0.01	-0.13	0.06	0.07	-0.17	-0.16	0.15
Cap	0.28	0.06	1.00	0.84	0.11	0.13	-0.15	0.33	-0.19	0.29	0.20	-0.17	-0.04	-0.19	0.05	-0.30	-0.19	-0.23	-0.18	0.13
LogC	0.25	0.04	0.84	1.00	-0.01	0.13	-0.12	0.28	-0.43	0.36	0.36	-0.20	0.02	-0.23	0.00	-0.41	-0.21	-0.22	-0.16	0.19
Group	0.13	-0.04	0.11	-0.01	1.00	-0.12	-0.02	-0.05	0.19	0.03	-0.17	0.14	0.07	-0.21	0.03	0.24	-0.12	-0.06	-0.05	0.02
Align	0.08	0.07	0.13	0.13	-0.12	1.00	-0.60	-0.11	0.00	0.08	0.13	-0.18	-0.23	-0.10	-0.01	-0.31	-0.18	-0.10	-0.07	0.09
TPC	-0.31	-0.04	-0.15	-0.12	-0.02	-0.60	1.00	-0.28	-0.17	-0.18	-0.01	0.05	0.30	0.31	-0.02	0.10	0.36	0.30	0.27	-0.24
Age	0.46	-0.07	0.33	0.28	-0.05	-0.11	-0.28	1.00	0.05	0.37	0.07	0.15	-0.14	-0.31	0.13	0.12	-0.43	-0.37	-0.34	0.40
LTOS	-0.08	-0.02	-0.19	-0.43	0.19	0.00	-0.17	0.05	1.00	0.01	0.08	0.30	0.05	-0.12	-0.04	0.30	-0.09	0.15	0.18	0.02
Invmt	0.39	0.14	0.29	0.36	0.03	0.08	-0.18	0.37	0.01	1.00	0.16	0.06	-0.14	-0.13	0.02	0.00	-0.18	-0.21	-0.16	0.33
RE1	-0.14	-0.07	0.20	0.36	-0.17	0.13	-0.01	0.07	0.08	0.16	1.00	0.37	0.05	-0.14	0.01	-0.21	-0.16	0.20	0.42	-0.12
RE2	-0.15	-0.03	-0.17	-0.20	0.14	-0.18	0.05	0.15	0.30	0.06	0.37	1.00	0.29	-0.24	-0.01	0.71	-0.07	0.16	0.39	-0.10
Acq	-0.39	-0.02	-0.04	0.02	0.07	-0.23	0.30	-0.14	0.05	-0.14	0.05	0.29	1.00	-0.15	-0.08	0.27	0.23	0.18	0.19	-0.11
Admin	-0.32	0.01	-0.19	-0.23	-0.21	-0.10	0.31	-0.31	-0.12	-0.13	-0.14	-0.24	-0.15	1.00	-0.04	-0.17	0.81	0.29	0.25	-0.26
Forex	0.15	-0.13	0.05	0.00	0.03	-0.01	-0.02	0.13	-0.04	0.02	0.01	-0.01	-0.08	-0.04	1.00	-0.03	-0.14	-0.15	-0.14	0.06
PrRet	0.03	0.06	-0.30	-0.41	0.24	-0.31	0.10	0.12	0.30	0.00	-0.21	0.71	0.27	-0.17	-0.03	1.00	0.07	-0.09	-0.03	0.07
WVsE	-0.34	0.07	-0.19	-0.21	-0.12	-0.18	0.36	-0.43	-0.09	-0.18	-0.16	-0.07	0.23	0.81	-0.14	0.07	1.00	0.21	0.21	-0.31
NLR	-0.84	-0.17	-0.23	-0.22	-0.06	-0.10	0.30	-0.37	0.15	-0.21	0.20	0.16	0.18	0.29	-0.15	-0.09	0.21	1.00	0.93	-0.47
GLR	-0.81	-0.16	-0.18	-0.16	-0.05	-0.07	0.27	-0.34	0.18	-0.16	0.42	0.39	0.19	0.25	-0.14	-0.03	0.21	0.93	1.00	-0.51
PPaid	0.42	0.15	0.13	0.19	0.02	0.09	-0.24	0.40	0.02	0.33	-0.12	-0.10	-0.11	-0.26	0.06	0.07	-0.31	-0.47	-0.51	1.00

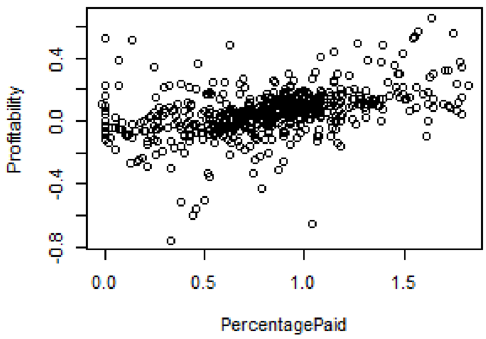


Figure 8 – Plot of Profitability versus PercentagePaid for the original dataset.

In order to get a better feeling for the association of two variables, it can also be helpful to look at two-dimensional plots in additon to the correlation coefficient. As an example, a plot of Profitability versus PercentagePaid is shown in Figure 8.

The relationship of each candidate predictor with the response variable Profitability is shown graphically in Figures 18, 19 and 20 in Appendix I (page 78).

However, an apparent association between a certain candidate predictor and Profitability cannot directly be interpreted as significant dependence because there might be correlations with other predictors as well. The graphs are thus just presented as a first visual impression of the dataset. It is through the linear model builing process that more insights into the associations of the different variables will be gained.

It should also be noted that for the purpose of plotting, certain data points were removed as described in the Appendix I.

Following these observations on the association of variables in the original dataset, we will now look at linear models that can be built from this data. Section 2.3 has recalled some theoretical concepts about linear models.

In this work, there were four different linear model building algorithms that were chosen to be explored with the aid of predefined R functions [9]. The corresponding R code is documented in Appendix I (page 71).

A fifth algorithm was somewhat different because it was based on a dynamic dataset. This will be described below.

Table 6 lists these approaches and indicates which criterion has been applied in which case.

9. Documentation about the application of such functions can be found for example in [21] and [19]

Table 6 – Approaches to linear model building and corresponding criteria that were chosen for this work. Details are given in the text. The labels a) to e) are used in the text.

Label	Approach	Criterion
a)	Forward selection	AIC
b)	Stepwise forward	AIC
c)	Stepwise backward	AIC
d)	Backward elimination	BIC
e)	Forward selection with dynamic dataset	t-test

In forward selection, candidate predictors are added in an iterative way in the aim of increasing the quality of the model according to the selected criterion. In backward elimination, candidate predictors are dropped one by one (starting from a model containing all of them).

The two stepwise methods allow for both adding and dropping of candidate predictors. The difference between the stepwise forward (b) and stepwise backward (c) approaches is that in the first case the first model under consideration does not contain any of the predictors (only the intercept) while in the second case it contains all of the candidate predictors.

As described in Section 2.3 in the Theoretical Part, the criteria AIC and BIC, which take into account the quality of fit and the number of retained predictors, penalize the number of retained predictors in a different way. BIC will favor models with a lower number of parameters than AIC.

The iterative exploration of the nested models can be imagined as a path on a multidimensional staircase which represents the value of the considered criterion as a function of whether or not each predictor is retained in the model. This path continues as long as it goes towards a minimum of the staircase, which is in the best case the global minimum, but might also be a local one. As mentioned in Section 2.3 in the Theoretical Part, the minimum has to be searched in a heuristic way because with 19 candidate predictors, there are too many possible models to test them all.

By exploring a longer path there might be more chance to find the global rather than a local minimum.

Thus, in the aim of maximizing the number of iterations when testing nested models, i.e. those moves which still optimize the criterion, we chose to use AIC for forward selection (a) and BIC for backward elimination (d).

For both of the stepwise algorithms, AIC was chosen in order to make them comparable to each other and to the approach a).

The different approaches a) to e) potentially lead to different results. As we will see later, in the case of the imputed data the paths taken by the different algorithms actually all end in the same minimum for a given criterion, suggesting that it is the global minimum.

In order to distinguish the results based on the original data that will be presented below from those based on imputed data that will be presented later, the numbers 1 and 2 will be used in the labels.

As an example, the output of 1)a) unimputed data forward selection is presented in Figure 9. It is one of the approaches that has not been very conclusive, meaning that the retained predictors and the coefficients are far from optimal, permitting us to display it as an example in the non-confidential part of this document.

The other outputs of linear modeling in R are given in the confidential Appendix II (page **??**).

The results off all approaches will be discussed in a comparative manner in Section 3.4.5 below.

In the output provided by R, the coefficients β can be seen in the first column. The last column displays the p-values from the individual t-tests. Each of these t-tests tests the two-sided hypothesis that the true coefficient of the corresponding predictor is equal to zero.

If this hypothesis is accepted (p-value above the level of significance), the implication is that no linear relationship exists between the response variable and the considered predictor. In contrast, if the p-value is below the level of significance, this hypothesis is rejected, meaning that the predictor is considered to significantly contribute to the linear model.

The example above shows that several predictors are retained as significant in the model. It can happen that the significance of a predictor is lost when other predictors are added. In the example above, this seems to have happened to NLR, probably at the stage where GLR was added (the two being correlated).

Overall, the linear models trained with the original data and the approaches a) to d) are not very convincing. It would be good to build linear models based on more data points.

This is why the approach e) was developed. By writing an algorithm for forward selection rather than using the automatic functions, it is possible to maximize the number of data points that are taken into account at each step. The corresponding R script is exhibited in Appendix I. In this approach, the data points (rows) with blanks are not removed a priori but simply left to be removed by the actual linear-model-function in the software R. This means that the different models that are investigated are based on different numbers of data points.

Figure 9 – Output of linear modeling in R of approach 1)a). The structural formula for this linear model (written without the coefficients because it is just supposed to show the structure of the model and it corresponds to the notation used in R) is: Profitability ÑLR + Acquisition + InvestmentReturn + Admin + Alignment + GLR + PercentagePaid + ReinsEff2 + PremiumRetention + GroupSupport

```
Residuals:
      Min         1Q     Median         3Q        Max
-0.149727  -0.032282  -0.001082   0.026811   0.283797

Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)       1.0119705  0.1389635   7.282 5.22e-11 ***
NLR              -0.0634766  0.1436225  -0.442  0.65938
Acquisition      -0.4586654  0.0704225  -6.513 2.28e-09 ***
InvestmentReturn  1.1739824  0.2242316   5.236 7.97e-07 ***
Admin            -0.0667416  0.0558080  -1.196  0.23430
Alignment        -0.0003593  0.0002344  -1.533  0.12813
GLR              -0.7851022  0.1929207  -4.070 8.89e-05 ***
PercentagePaid   -0.0274488  0.0114161  -2.404  0.01787 *
ReinsEff2         0.7114507  0.2123730   3.350  0.00111 **
PremiumRetention -0.3321710  0.1110416  -2.991  0.00343 **
GroupSupport      0.0003674  0.0001809   2.032  0.04461 *
---
Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05683 on 110 degrees of freedom
Multiple R-squared:  0.8524,    Adjusted R-squared:  0.839
F-statistic: 63.53 on 10 and 110 DF,  p-value: < 2.2e-16
```

Here, the p-value for the t-test for the tentatively added coefficient is used as a criterion for forward variable selection. The lowest p-value is identified and if it is below the chosen level of significance (here $\alpha = 5\%$), the corresponding variable is added to the model. When adding or dropping a single coefficient from the model, the individual t-test is equivalent to the F-test for comparing the models.

The linear model obtained by the approach 1)e) has the structure show in Figure 10.

It can be observed that many predictors do no longer have significant p-values in the final model even though each p-value was significant at least in the iteration when that predictor was selected to be included in the model.

Figure 10 – Output of linear modeling in R of approach 1)e).

```
Residuals:
      Min         1Q     Median         3Q        Max
-0.262242  -0.025685   0.000563   0.025549   0.271856

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.7953627  0.1489880   5.338 2.86e-07 ***
GLR              -0.5911443  0.1938017  -3.050 0.002640 **
NLR              -0.1215973  0.1562989  -0.778 0.437626
Acquisition      -0.4910071  0.0711804  -6.898 9.13e-11 ***
Age               0.0011291  0.0007686   1.469 0.143613
ReinsEff2         0.4591721  0.2274528   2.019 0.045030 *
WrittenVsEarned   0.0786151  0.0201788   3.896 0.000139 ***
BasisMarket      -0.0665961  0.1652589  -0.403 0.687451
PremiumRetention -0.2351237  0.1237874  -1.899 0.059145 .
ReinsEff1        -0.0045994  0.0256048  -0.180 0.857651
LogCapacity       0.0038879  0.0122244   0.318 0.750830
GroupSupport      0.0004536  0.0001514   2.996 0.003126 **
InvestmentReturn  1.2010327  0.1486125   8.082 9.86e-14 ***
Admin            -0.3099650  0.0735422  -4.215 3.99e-05 ***
TPC               0.0009767  0.0004259   2.293 0.023005 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06209 on 176 degrees of freedom
  (429 observations deleted due to missingness)
Multiple R-squared:  0.8229,    Adjusted R-squared:  0.8088
F-statistic: 58.41 on 14 and 176 DF,  p-value: < 2.2e-16
```

Even though this is not ideal, it is what is often observed when working with real-life data. The iterative approach chosen here is just one of several possible paths. In order to be sure that linear modeling is more conclusive, one would have to look at all the possible sub-models and choose the best one for a given criterion, but as described in Section 2.3, this is not practically possible with 19 predictors.

3.4.3 Models trained with imputed data

The majority of the missing data concerns the predictors Alignment, Group-Support and LTOS. That is why the data in these columns was imputed in the following manner:

- The Alignment value for all the 6xxx syndicates, i.e. the SPAs, was set to 100% for all years (no data given in SRL) [10].
- For both GroupSupport and Alignment, the blanks for 2013 and earlier years were filled with the average of the 2016, 2015 and 2014 values of the corresponding syndicate.
- The 145 blanks in the column LTOS were all filled with the average of all existing datapoints in that column, namely 0.4613.

This imputation resulted in 463 complete data points. The imputed data was then used to build linear models with the same procedure as described above for the original data.

Again, we will start by discussing the correlation matrix of the dataset which is displayed in Figure 7.

The Pearson correlation matrix for the imputed data as displayed in Table 7 shows the following similarities with the correlation matrix for the data before imputation:

- As above, LogCapacity and Capacity are correlated with ($r = 0.82$ here, $r = 0.84$ above).
- The negative correlation between Alignment and TPC is still present even though it is now just -0.49 (and thus not highlighted in color).
- The dependence of Profitability on NLR and GLR is less pronounced than above (-0.60 and -0.66 compared to -0.84 and -0.81 respectively). This means that across the market cycle, there are more differences between loss ratios and profitability than in the years 2014-2016. Possible explanations could include the trends and patterns in investment returns, acquisition or administrative expenses.
- The associations between ReinsEff2 and PremiumRetention, between WrittenVsEarned and Admin as well as between GLR and PercentagePaid as discussed above are still present to a similar extent.

10. This is a simplifying assumption because there are SPAs with third party participation.

Table 7 – Based on the imputed ($n = 463$ complete data points): Pearson correlation matrix for the dependent variable Profitability and the 19 independent variables. Coefficients above 0.5 are highlighted in red, those below -0.5 in blue. Prof=Profitability, BM=BasisMarket, Cap=Capacity, LogCap=LogCapacity, Group=GroupSupport, Align=Alignment, Invmt=InvestmentReturn, RE1(2)=ReinsEfficiency1(2), Acq=Acquisition, PrRet=PremiumRetention, WVsE=WrittenVsEarned, PPaid=PercentagePaid.

	Prof	BM	Cap	LogC	Group	Align	TPC	Age	LTOS	Invmt	RE1	RE2	Acq	Admin	Forex	PrRet	WVsE	NLR	GLR	PPaid
Prof	1.00	0.28	0.17	0.16	0.10	-0.01	-0.08	0.28	-0.07	0.40	-0.20	-0.18	-0.21	-0.03	0.10	0.02	-0.18	-0.60	-0.66	0.11
BM	0.28	1.00	-0.03	-0.04	0.00	0.00	-0.03	-0.05	-0.02	0.13	-0.19	-0.22	-0.02	-0.01	-0.09	0.00	0.04	-0.25	-0.30	0.14
Cap	0.17	-0.03	1.00	0.82	-0.04	0.09	-0.12	0.32	-0.13	0.13	0.09	-0.11	0.02	-0.03	0.01	-0.28	-0.15	-0.06	-0.06	0.03
LogC	0.16	-0.04	0.82	1.00	-0.11	0.05	-0.10	0.32	-0.35	0.20	0.19	-0.09	0.03	-0.08	0.02	-0.35	-0.26	-0.03	-0.01	0.00
Group	0.10	0.00	-0.04	-0.11	1.00	-0.09	0.02	-0.16	0.16	-0.06	0.04	0.12	-0.04	-0.04	0.02	0.21	-0.04	-0.02	-0.01	-0.07
Align	-0.01	0.00	0.09	0.05	-0.09	1.00	-0.49	-0.22	0.04	0.06	0.00	-0.04	0.02	0.15	0.04	-0.17	0.08	0.09	0.06	0.02
TPC	-0.08	-0.03	-0.12	-0.10	0.02	-0.49	1.00	-0.15	-0.20	-0.11	-0.05	-0.03	-0.05	-0.17	-0.05	0.06	0.04	-0.07	-0.02	-0.06
Age	0.28	-0.05	0.32	0.32	-0.16	-0.22	-0.15	1.00	-0.03	0.21	0.10	0.07	-0.01	-0.01	0.00	-0.03	-0.28	-0.16	-0.16	0.15
LTOS	-0.07	-0.02	-0.13	-0.35	0.16	0.04	-0.20	-0.03	1.00	-0.01	0.07	0.16	0.07	0.04	-0.03	0.19	0.02	0.02	0.06	0.04
Invmt	0.40	0.13	0.13	0.20	-0.06	0.06	-0.11	0.21	-0.01	1.00	0.15	0.06	0.04	0.00	0.09	-0.15	-0.11	-0.07	0.02	0.18
RE1	-0.20	-0.19	0.09	0.19	0.04	0.00	-0.05	0.10	0.07	0.15	1.00	0.65	0.07	0.01	0.04	-0.10	-0.08	0.26	0.55	0.03
RE2	-0.18	-0.22	-0.11	-0.09	0.12	-0.04	-0.03	0.07	0.16	0.06	0.65	1.00	0.23	0.21	0.06	0.51	0.02	0.36	0.57	-0.01
Acq	-0.21	-0.02	0.02	0.03	-0.04	0.02	-0.05	-0.01	0.07	0.04	0.07	0.23	1.00	0.56	0.00	0.10	0.35	0.19	0.20	0.02
Admin	-0.03	-0.01	-0.03	-0.08	-0.04	0.15	-0.17	-0.01	0.04	0.00	0.01	0.21	0.56	1.00	0.00	0.16	0.65	0.44	0.30	-0.02
Forex	0.10	-0.09	0.01	0.02	0.02	0.04	-0.05	0.00	-0.03	0.09	0.04	0.06	0.00	0.00	1.00	-0.02	-0.03	0.00	0.03	0.00
PrRet	0.02	0.00	-0.28	-0.35	0.21	-0.17	0.06	-0.03	0.19	-0.15	-0.10	0.51	0.10	0.16	-0.02	1.00	0.10	0.15	0.03	-0.05
WVsE	-0.18	0.04	-0.15	-0.26	-0.04	0.08	0.04	-0.28	0.02	-0.11	-0.08	0.02	0.35	0.65	-0.03	0.10	1.00	0.34	0.28	-0.08
NLR	-0.60	-0.25	-0.06	-0.03	-0.02	0.09	-0.07	-0.16	0.02	-0.07	0.26	0.36	0.19	0.44	0.00	0.15	0.34	1.00	0.86	-0.21
GLR	-0.66	-0.30	-0.06	-0.01	-0.01	0.06	-0.02	-0.16	0.06	0.02	0.55	0.57	0.20	0.30	0.03	0.03	0.28	0.86	1.00	-0.18
PPaid	0.11	0.14	0.03	0.00	-0.07	0.02	-0.06	0.15	0.04	0.18	0.03	-0.01	0.02	-0.02	0.00	-0.05	-0.08	-0.21	-0.18	1.00

There are also some correlations that become apparent in the imputed dataset while they were not yet present in the low number of data points (and covered years) before imputation:

— The two measures for reinsurance efficiency are correlated with each other (0.65), corroborating the idea that they measure similar characteristics all while being different.
— Both of the reinsurance efficiencies are somewhat positively correlated with GLR. This suggests that the reinsurance program of syndicates works rather well when losses are high, which is expected.
— Admin and Acquisition expenses are associated. This is quite surprising. As it is not clear what this association might be related to, both variables are kept in the candidate predictors to get a chance to enter the linear model.

The results of approaches 2)a) to 2)e) are presented in Appendix II and will be discussed in a comparative manner in Section 3.4.5 below.

3.4.4 Exploring the forced exclusion of variables

As a further exploration in the linear model building process, it was tested how the application of additional constraints changes the outcome. The idea was to exclude certain variables from the dataset because their absence makes the model more relevant for practice or because they were already observed to be irrelevant.

Specifically, variables were excluded as follows:

— BasisMarket was removed because it is not a syndicate characteristic. It had anyways only been introduced to potentially explain some unexplained variation in the data, but it was found not to be useful.
— Capacity war eliminated because Log(Capacity) had been found to be more significant.
— NLR and GLR were excluded because in practice, they would not be known earlier than the actual profitability. When considering historical data, they are available and are helpful in predicting historical profitability, but if we plan to predict future profitability we should ideally be able to do it without relying on loss ratios.
— ReinsEff2 was excluded because ReinsEff1 is the more relevant definition in combination with the PremiumRetention variable.

This dataset will be referred to with the number 3. The algorithm e) (see Table 6) was applied.

The resulting linear model is shown in Figure 11. The model was quite successful in so far that most of the variables were still significant at the end of the forward selection process (unlike in 1)e) discussed on page 41). It is also notable that the variables GroupSupport and Age appear in this model while they are less present in other models. This phenomenon is especially pronounced for Age. It seems that Age, which is correlated to Profitability with $r = 0.28$ (see Table 7) becomes a crucial factor in predicting Profitability once we cannot rely on loss ratios.

Beyond these observations, the model based on dataset 3 was not further used in the process of establishing a syndicate segmentation method. This is related to the conclusion drawn from approach 1)e) according to which forward selection based on p-values is not an ideal algorithm in this context, see 46 below.

Figure 11 – Output of linear modeling in R of approach 3)e).

```
Residuals:
     Min       1Q    Median       3Q      Max
-0.69819  -0.05382  0.00467  0.05862  0.44667

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.0454757  0.0671411  -0.677 0.498540
Forex              0.4613794  0.2826847   1.632 0.103324
InvestmentReturn   1.2698296  0.1263735  10.048  < 2e-16 ***
PremiumRetention   0.0978242  0.0644056   1.519 0.129469
ReinsEff1         -0.1089917  0.0134165  -8.124 4.06e-15 ***
GroupSupport       0.0007781  0.0001898   4.100 4.87e-05 ***
Age                0.0036218  0.0009181   3.945 9.21e-05 ***
Acquisition       -0.3740694  0.0609528  -6.137 1.79e-09 ***
LogCapacity        0.0357611  0.0128895   2.774 0.005751 **
Admin              0.0717887  0.0190273   3.773 0.000182 ***
WrittenVsEarned   -0.0404313  0.0185800  -2.176 0.030050 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1253 on 468 degrees of freedom
  (141 observations deleted due to missingness)
Multiple R-squared:  0.3757,    Adjusted R-squared:  0.3624
F-statistic: 28.17 on 10 and 468 DF,  p-value: < 2.2e-16
```

3.4.5 Conclusions from the linear models for profitability

As stated at the beginning of Section 3.4, the motivation for building linear models was to increase the understanding of interdependencies between variables and to select relevant variables for the development of a segmentation method.

By looking at the outputs from the different algorithms, we get the impression that some predictors are retained more frequently than others and with lower p-values. Figure 12 shows a graphical summary of the p-values of the predictors found in the different linear models.

The values of the coefficients in the various linear models are given in Appendix II for confidentiality reasons. The corresponding discussion concludes that there are some predictors that are consistently appearing with about the same coefficients across the different models while other predictors have coefficients are subject to large variations, which suggests interferences of these predictors with other predictors.

By comparing AIC values across models trained with the same dataset, it can be judged which one is the best according to AIC. Among 1)a), 1)b) and 1)c), the AIC is continuously decreasing, meaning that the algorithm c) stepwise backward was most successful in minimizing AIC for models based on the dataset 1 (unimputed data).

As becomes clear from the colored table in Figure 12, the unimputed dataset is less suitable than the imputed dataset for finding a model with clearly significant predictors. This is no surprise given the big difference in the number of data points.

Therefore, we will now focus solely on the models trained with the imputed dataset.

The methods 2)a), 2)b) and 2)c) actually converge to the same model. It thus seems that all of these algorithms are successful at finding the model that minimizes AIC over all possible models. To further confirm this finding, it was also tested which model is suggested by the backward elimination algorithm with AIC as decision criterion (not listed in Table 6). It was indeed once again the same model that was found.

The same holds true for the method 2)d) if it is complemented by the methods forward selection, stepwise forward and stepwise backward with BIC as a decision criterion (methods that are not listed in Table 6).

Given that AIC and BIC are two decision criteria among which there is no clear best choice [18], both models found with them are further considered. They form the basis for the development of a syndicate segmentation method as described in the section "LM" in Appendix II (page **??**).

When comparing 1)a) to 1)e), two forward selection algorithms, it can be observed that the change from AIC to p-values as decision criterion has made the model less conclusive. The model resulting from 1)e) does actually contain several variables that were significant at some point of the iteration but lost their significance through the addition of other variables to the model. Given that 1)a) and 1)e) also differ in the dataset, with 1)e) using a dynamic dataset which is on average larger than the dataset for 1)a) and should thus represent an advantage, the forward selection based on p-values seems even successful than the forward selection based on AIC.

The comparison of 2)e) to the other models based on the dataset 2 (the imputed data) shows only minor differences. This is a further confirmation of the consistency of these models.

As for the model resulting from 3)e), as described above (page 45), the apparition of GroupSupport and Age in the retained predictors is noteworthy, but the model was not considered conclusive enough for further use.

	original dataset				imputed dataset				blanks not removed a priori, i.e. dynamic dataset		
	fwd sel 1)a)	step fwd 1)b)	step bwd 1)c)	bwd elim 1)d)	fwd sel 2)a)	step fwd 2)b)	step bwd 2)c)	bwd elim 2)d)	1)e)	2)e)	3)e)
INTERCEPT	5E-11	2E-16	2E-16	2E-16	2E-16	2E-16	2E-16	2E-16	3E-07	8E-13	5E-01
BasisMarket					3E-02	3E-02	3E-02		7E-01	2E-02	EXCLUDED
Capacity											EXCLUDED
LogCapacity					6E-07	6E-07	6E-07	4E-06	8E-01	3E-05	6E-03
GroupSupport	4E-02	2E-02	3E-02	2E-02	7E-05	7E-05	7E-05	7E-05	3E-03	1E-05	5E-05
Alignment	1E-01	2E-01			2E-01	2E-01	2E-01				
TPC			3E-02						2E-02		
Age			8E-02						1E-01	1E-01	9E-05
LTOS			1E-01								
InvestmentReturn	8E-07	7E-07	5E-06	7E-07	2E-16	2E-16	2E-16	2E-16	1E-13	2E-16	2E-16
ReinsEff1									9E-01	6E-01	4E-15
ReinsEff2	1E-03	7E-11	4E-08	1E-10	2E-16	2E-16	2E-16	2E-16	5E-02	6E-09	EXCLUDED
Acquisition	2E-09	3E-09	4E-09	6E-09	2E-16	2E-16	2E-16	2E-16	9E-11	2E-16	2E-09
Admin	2E-01		3E-02		2E-16	2E-16	2E-16	2E-16	4E-05	4E-15	2E-04
Forex					3E-03	3E-03	3E-03	8E-03		4E-03	1E-01
PremiumRetention	3E-03	6E-06	7E-06	2E-05	1E-04	1E-04	1E-04	7E-04	6E-02	2E-02	1E-01
WrittenVsEarned			6E-02		1E-01	1E-01	1E-01		1E-04	1E-01	3E-02
NLR	7E-01								4E-01	8E-01	EXCLUDED
GLR	9E-05	2E-16	2E-16	2E-16	2E-16	2E-16	2E-16	2E-16	3E-03	2E-16	EXCLUDED
PercentagePaid	2E-02	2E-02	2E-02	1E-02	1E-06	1E-06	1E-06	3E-06		2E-06	

Figure 12 – Overview of the p-values for different linear models. The detailed explanation of the model building approaches 1)a) to 3)e) is given in the text. The color scale goes from green for the smallest values to red for the largest values and serves as a guidance for the eye. Blank cells correspond to predictors that were not retained in the respective model.

The following list summarizes the predictors that were found to be most relevant [11] in those two models:

- LogCapacity
- InvestmentReturn
- ReinsEff2
- Acquisition
- Admin
- GLR
- PercentagePaid

This means that these factors deserve particular attention when developing methods for syndicate segmentation.

The coefficients for these predictors as obtained from the conclusions of linear model building are presented in Appendix II. They served as a basis for a formula to predict the profitability of a syndicate, see Section 3.5.

Before moving on to actually developing and testing the segmentation methodologies, we can discuss some limitations of the approach to predict profitability from linear model building. There are a number of general things that should be kept in mind when working with stepwise regression procedures [22] [20]:

- The search algorithms are not guaranteed to find the best model according to the chosen criterion, so the final models are not guaranteed to be optimal in any specific sense.
- There are often numerous equally good models, but only a single one of them is appearing as the final model for any given search algorithm and criterion.
- Neither the presence/absence of certain predictors nor the order in which the predictors enter the model should be over-interpreted. It is possible that unimportant predictors have not been eliminated or that not all of the important predictors have been identified.
- Existing knowledge is not taken into account in the iterative regression, so it might be necessary to adapt the procedure in order to include important predictors [12].
- As in any statistical analysis, it is possible to commit Type I or Type II errors. It should in particular be kept in mind that during an iterative regression procedure, a high number of t-tests for testing $\beta_k = 0$ are conducted, meaning that it is quite likely that in some of them errors were committed.

11. p-value below 10^{-5} used as decision criterion.
12. This was attempted with approach 3)e) above.

Therefore, in conclusion, the results of the linear model building process should not be over-interpreted.

3.5 Selection of criteria and weighting for segmentation

There are two methodologies for syndicate segmentation that have been developed in this work.

As described above, we can imagine to segment syndicates by ranking them according to the profitability predicted by a linear model.

After having chosen a linear model, it has to be decided how many years of data are taken into account to predict future profitability. We decided to take into account the latest 8 CALs, i.e. the same period as used for the building of the linear models.

This implies of course that we are not really predicting *future* profitability, but the profitability observed over the last 8 years. However, the recent past is still the best indication that we have for the future. The scoring systems of SRL and APCL are also mainly based on past profitability (see Section 2.2.1).

An alternative method for syndicate segmentation has already been developed in PartnerRe's "Capital at Lloyd's" team based on professional intuition. A scoring method for five quantitative factors as well as a weighting rule was defined in order to obtain an overall score. This was then combined with external ratings and some additional factors were used in a binary way to include or exclude syndicates from the ranking.

These two methodologies are described in Appendix II (page **??**) for confidentiality reasons and labeled as "LM" and "PRE" respectively.

3.6 Main results and discussion

This Section presents a comparison of the different segmentation methods in action. In addition, a few observations about benefits and costs as well as some ideas for further improvement of the methods are exposed.

3.6.1 Comparing the performance of portfolios selected with different methods

In order to quantify the value created by the new syndicate segmentation methodologies, we will compare the performance of portfolios of syndicates generated with several different segmentation approaches.

Given that the goal of PartnerRe's "Capital at Lloyd's" team is to build a portfolio of about 15 syndicates, the top 15 syndicates were determined with each method.

It is clear that a portfolio of exactly those selected syndicates cannot be realized in practice because it would include syndicates which do not take TPC. Nevertheless, this type of hypothetical portfolio was considered as the most suitable one for comparison purposes. Some aspects of the selection of target syndicates in practice will be discussed in Section 3.7.

Here, the different rankings from which we took the top 15 syndicates were established in the following ways:

— "LM": based on the predicted profitability using a formula obtained from linear modeling (see Section "LM" in Appendix II)
— "PRE": based on the segmentation framework that is based on professional experience and accounts for some of PartnerRe's strategic preferences (see Section "PRE" in Appendix II)
— "SRL": based on the SRL ratings
— "APCL": based on the APCL ratings
— "Random": completely randomly (each of the 96 syndicates with an equal chance of selection)

The top 15 lists obtained with these approaches are displayed in Table **??** in Appendix II.

For the comparison of these ranking methods, the considered performance measure is profitability - but which one exactly?

One idea would be to look at the profitability of the latest YOA.

However, the time horizon of interest for PartnerRe as a capital provider in the Lloyd's market is much longer. In practical terms, it is the ROC over any 3-5 year period that matters to the company. That is why the average profitability of the latest 4 YOAs is considered as the most relevant performance measure. In the graphs presenting the results, this will be abbreviated as "4YOA". It should however be kept in mind that the overlap between the data for this performance measure (YOAs 2013-2016, including forecasts for 2015 and 2016) and the data used to train the linear models (CALs 2009-2016) is quite large even though they do not come from the same accounting system.

To complement this performance measure with one that takes into account as much data as possible, the long-term average ROC is also compared across the selected portfolios. It is defined at the simple average of rolling 5-CAL-averages of ROC [13] over the period 2004-2016. It is labeled "LTROC" in the graphs. For those syndicates that are comparatively old, there is thus some data in this dataset that was not used to train the linear models.

13. Obtained from profitability and the assumption that capital is 60% of capacity introduced in Section 2.1.4.

On the other hand, it could also be interesting to look at the performance that the syndicates achieved in the latest closed YOA, namely 2014. This performance measure will be referred to as "2014 YOA" in the labels in the plots. It has the advantage of reducing the overlap of the training and the test datasets to a minimum [14].

In summary, we will look at three measures of profitability:

— "4YOA"
— "LTROC"
— "2014 YOA"

It should be noted that for the segmentation of syndicates based on the SRL and APCL ratings, it was necessary to define an additional criterion because there are many syndicates with the same rating. The choice fell on the "4YOA" profitability of syndicates, so syndicates are first sorted by decreasing "4YOA" profitability and then sorted by the SRL or APCL rating. It has to be kept in mind that this creates an artificial advantage for these methods when looking at the corresponding "4YOA" portfolio profitability. However, the curves of portfolio profitability as a function of the number of syndicates is anyways discussed in a holistic way rather than point by point, which decreases the importance of the actual ranks.

As stated above, the top 15 syndicates were determined with each method. However, it is not enough to select the syndicates, but it must also be determined which share of capacity is taken on each syndicate.

One approach would be to take a certain amount of capacity on each syndicate, for example 20 GBPm, no matter the size of the syndicate.

In contrast, another approach could be to always try to take a certain share of a syndicate's capacity, for example 15%. This would however be subject to an upper limit of line size, for example 50 GBPm because 15% of a very big syndicate would be undesirably large. Hence, a mathematical rule for determining the share on each syndicate could be $\min(15\%\text{Capacity}; 50\text{ GBPm})$.

In practice, there are no strict rules as to how PartnerRe's "Capital at Lloyd's" team determines the amount of capacity taken on each syndicate. Much of it is a matter of negotiation and depends on the context of the investment.

In summary, we will look at two portfolio compositions:

— Construction rule 1: Equal line size on each syndicate

14. In order to avoid any overlap, it would be imaginable to use for example a k-fold cross-validation approach, i.e. randomly partition the sample in k subsets, train the model with $k-1$ of those and test it with the k-th subset, averaging the results over the k repetitions of this process.

— Construction rule 2: Taking a 15% share of each syndicate but no more than 50 GBPm

Figures 13, 14 and 15 present the three performance measures of these various portfolios built with the two construction rules and with the four syndicate segmentation methods that should be compared to each other (five including "Random").

A number of points should be noted about these graphs. First of all, what is plotted is the average profitability of the portfolio as a function of the number of syndicates in the portfolio. This yields a **decreasing curve** because an ideal method ranks the syndicates such that the first one has the highest profitability.

However, this does not imply that one should participate only in a low number of syndicates because the absolute profit in GBP is still increasing when the next syndicates are added to the portfolio (see the alternative plot in Figure 16). We thus face a trade-off between total portfolio size (total capacity) and average profitability.

A second point that should be kept in mind when interpreting these graphs concerns the handling of **missing data**. The syndicates for which data for the considered performance measure is missing are nevertheless kept in the respective portfolio. It is assumed that their performance equals the average performance of all the syndicates that were added to the portfolio previously, yielding a flat segment of the curve.

This assumption is introducing a positive bias, but it is acceptable given that the number of missing data points is similar in the portfolios that should be compared to each other (except for "APCL" where no data is missing but which is anyways rather better than the others because of its bias for syndicates with TPC).

In contrast, the alternative plot in Figure 16 represents portfolios which contain only syndicates for which the required data is available. This introduces a bias as well because, depending on the positions at which discarded syndicates appear in a top 15 list, the curves are shifted horizontally in unequal ways.

Another point that should be noted about the chosen graphical representation is that the **curve for the random selection** of syndicates is only partly meaningful because the order of syndicates is interchangeable there. However, by chance the first random number corresponded to a syndicate with a high profitability, meaning that the next syndicates pull down the average, thereby yielding a curve similar to those ones based on the actual segmentation methods.

Finally, between **the profitability and the ROC graphs**, there is an intrinsic difference of a factor 1.67 because the ROC is calculated based on the assumption that capital is 60% of capacity, see Section 2.1.4.

Let's start by discussing the trends observed in the **"4YOA" graphs**:

- All methods are much better than a random selection of syndicates, which underlines the interest of doing some efforts in syndicate segmentation. It should be noted that even the random selection is expected to be better than what a capital provider would get if (s)he were just taking the offers made by syndicates looking for capital. This is because it is generally the distressed or the start-up syndicates which are searching for new capital providers. Therefore it is necessary to be attentive not to be selected against. A segmentation method can definitely help in this respect.
- The methods "PRE" and "SRL" yield quite similar curves. This is not surprising knowing that the corresponding top 15 rankings are similar, see Table **??** in Appendix II. This stems from the fact that the underlying methodologies present also quite some overlap (see Section 2.2.1 and the methodology description "PRE" in the confidential Appendix II (page **??**) respectively).
- "LM" is the method that achieves the highest profitability for a very small portfolio. However, it should be noted that these values depend very much on the factors that caused the exclusion of certain syndicates in certain methods. The "SRL" method cannot suggest syndicates for which no SCO is assigned by SRL. The "PRE" method works with certain pre-defined exclusion criteria. And the "LM" method is also a combination of a formula and a filtering step as described in the section "LM" in Appendix II. Overall, not too much attention should be paid to the individual points of the plots.
- The "APCL" method has resulted in a portfolio in which the most profitable syndicates are only added in the 6th to 8th position, which is due to the fact that they got only the second-highest of the actually attributed scores in the APCL rating. For a portfolio comprising 8 to 15 syndicates, the "APCL" method yields comparable profitability levels as the "LM" method.
- When comparing the two graphs corresponding to the two portfolio construction rules, it can be noted that for the first one (equal line size), the "LM" and "APCL" methods are the best. This is because in those top 15 lists, but not in those based on the "PRE" and "SRL" methods, we can find some highly profitable syndicates that are rather small. Such syndicates contribute to a high average profitability all over the portfolio, but their effect vanishes in case of the second portfolio con-

struction rule (15% share, ≤50 GBPm) where larger syndicates enter with much more weight (up to a limit of 50 GBPm). The small absolute contribution to portfolio profit of the syndicates that achieve the highest ranks in the "LM" and "APCL" methods can also be observed in the alternative plot in Figure 16. Depending on the specific portfolio building context, different methods have different advantages [15].

When looking at the **"LTROC" graphs**, the following additional observations can be made:

— The syndicates selected as numbers 10 to 15 in the "APCL" method were apparently less profitable in the distant past than in the recent past.

— The portfolios built with the "LM" method are the best in both cases, not only in the case of portfolio construction rule 1 as above.

Finally, the **plots for the "2014 YOA"** reveal some further specificities:

— The syndicate ranked second in the "LM" method and 6th in the "APCL" method was by chance very highly profitable in the "2014 YOA" and thus distorts the curves. It is one of the syndicates that would most certainly not become part of PartnerRe's portfolio due to other considerations (see Section 3.7 about syndicate selection in practice).

— Overall, similar observations as in "4YOA" can be made, namely the conclusion that for a portfolio of 10 to 15 syndicates, the methods "LM", "PRE", "SRL" and "APCL" all yield quite similar results which are much better than for a random selection.

3.6.2 Comparing the benefits and costs of the different methods

As discussed in the last section, the profitability of a portfolio can be increased if one of the segmentation methods is applied instead of selecting the syndicates randomly (or even worse, according to the most apparent demand for capital).

The benefit of these methods is thus a higher ROC, even though the actual benefit in practice is difficult to quantify (because the analysis here is based on many assumptions and does not account for the practical aspects of pursuing investment opportunities).

15. It was also explored how the different methods perform in the case of portfolio construction rules that are a variation of the construction rule 2. When the maximum line size is kept small, the "LM" and the "APCL" segmentation approaches outperform the two others like in the equal line size portfolio (because a very small maximum line size is achieved in all syndicates). With a maximum line size of some 20 GBPm or more, the difference between the methods vanishes for the range of 8 to 15 syndicates.

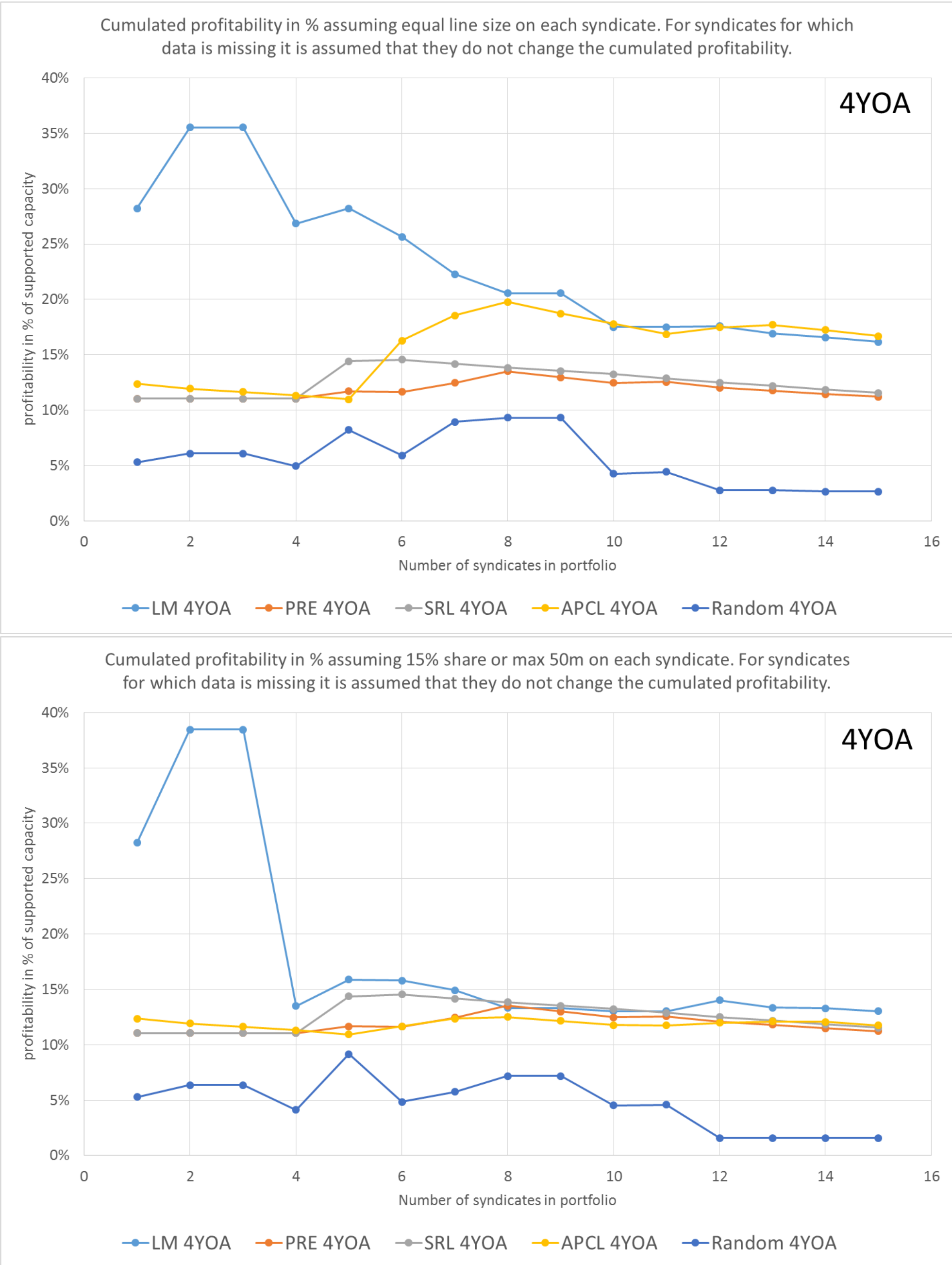
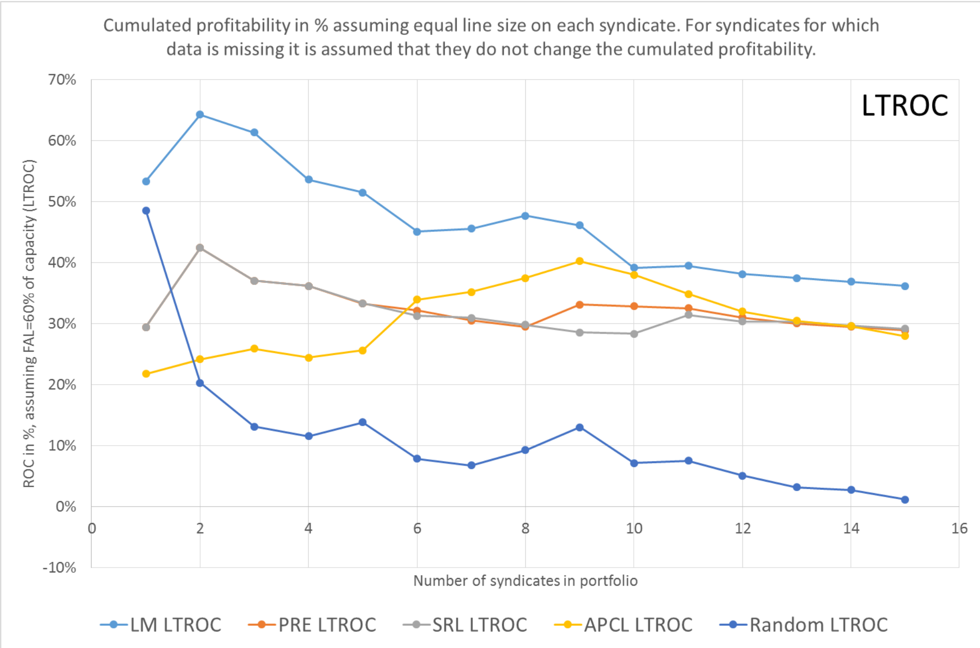


Figure 13 – Comparison of the "4YOA" performance of portfolios built by different methods. "LM", "PRE", "SRL", "APCL" and "Random" designate the different segmentation methods (see page 51). The upper graph shows the portfolios constructed with the portfolio construction rule 1, the lower graph with construction rule 2 (see page 52).



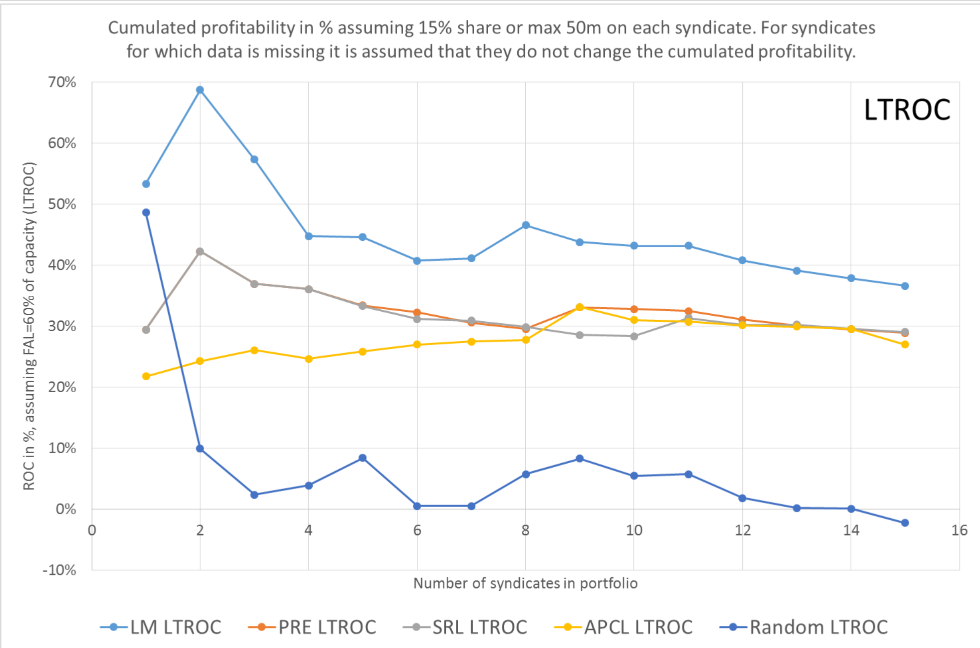
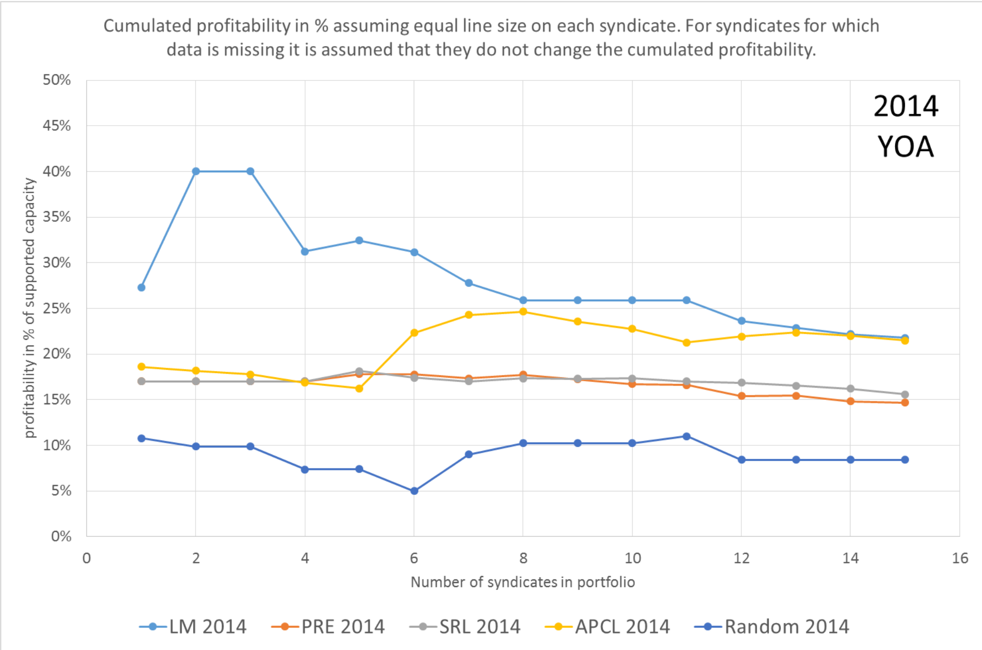


Figure 14 – Comparison of the "LTROC" performance of portfolios built by different methods. "LM", "PRE", "SRL", "APCL" and "Random" designate the different segmentation methods (see page 51). The upper graph shows the portfolios constructed with the portfolio construction rule 1, the lower graph with construction rule 2 (see page 52).



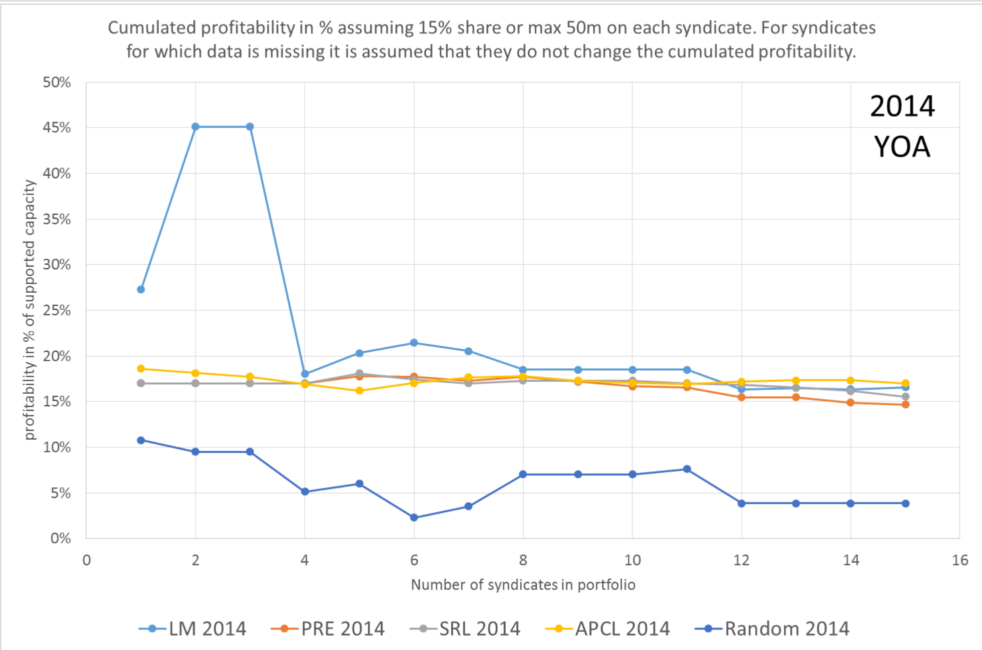
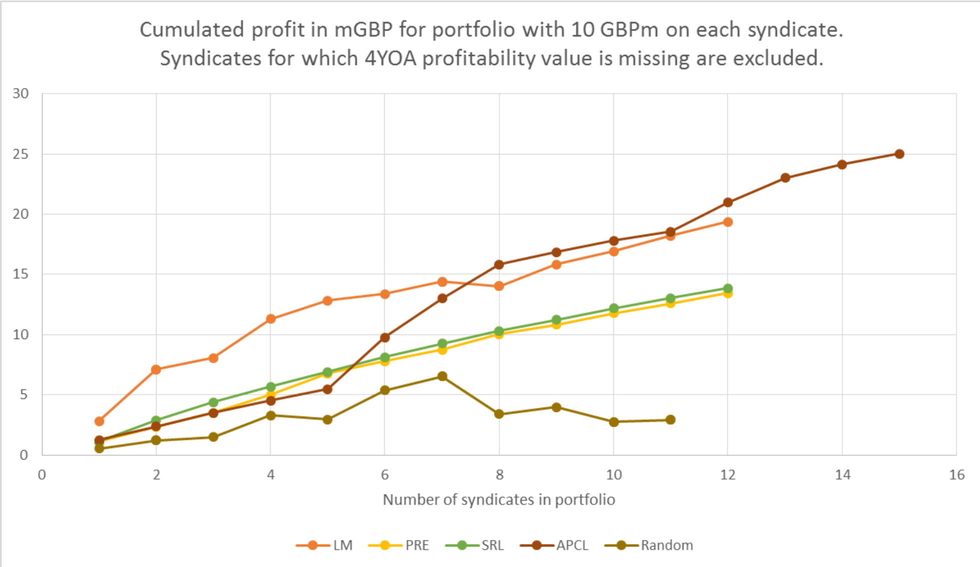


Figure 15 – Comparison of the "2014YOA" performance of portfolios built by different methods. "LM", "PRE", "SRL", "APCL" and "Random" designate the different segmentation methods (see page 51). The upper graph shows the portfolios constructed with the portfolio construction rule 1, the lower graph with construction rule 2 (see page 52).



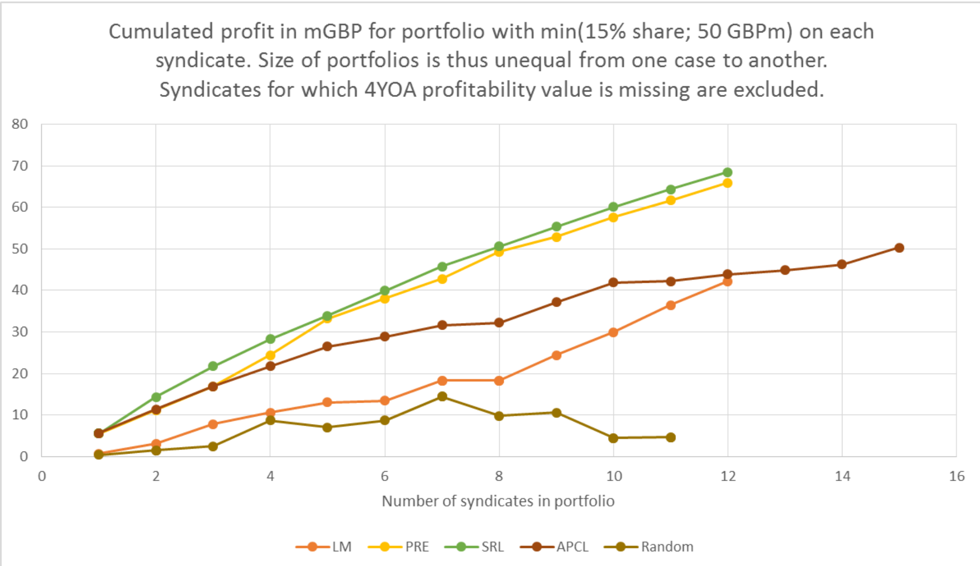


Figure 16 – Alternative graphical representation of the comparison of the "4YOA" performance of portfolios built by different methods. "LM", "PRE", "SRL", "APCL" and "Random" designate the different segmentation methods (see page 51). The upper graph shows the portfolios constructed with the portfolio construction rule 1, the lower graph with construction rule 2 (see page 52).

The segmentation method based on insights from linear modeling outperforms the other methods in some specific cases. However, this should not be overstated due to the partial overlap of the training and test datasets and due to all kinds of other limitations of the methodologies.

Overall, the rankings obtained by the different methods are similar. There are some syndicates that are consistently identified as attractive investment opportunities.

This work has thus allowed to confirm the accuracy of the SRL and APCL rankings in an independent manner. When looking at those rankings in the future, their predictions can be used with an increased confidence.

In terms of costs, the different methods are quite comparable. The subscription to SRL's database and to the Lloyd's Statistics as well as the fees paid to the Members' Agent are fixed expenses. Once the data is obtained, the different methods for making use of them are similar in terms of time consumption. Especially now that the methods are already defined, implemented and compared, the main efforts in the future will be to keep PartnerRe's internal database updated and to interpret the newly generated rankings.

3.6.3 Ideas to further improve the segmentation methods

There are a variety of potential improvements that could be made to the segmentation methods. In general, the process of refining the understanding of syndicate dynamics by statistical modeling could be taken much further. By combining even more data, the models could be refined and the multi-criteria segmentation approaches could become even more nuanced.

First of all, the data that will be accumulated over the years of activity of PartnerRe's "Capital at Lloyd's" team will provide a basis for more refined syndicate segmentation.

Second, additional data sources might open new possibilities for taking into account additional syndicate characteristics. For example, if RDS percentages could be obtained for the non-APCL syndicates as well, it would be interesting to make use of them in the segmentation method. This also relates to the separate treatment of profitability and volatility as discussed in Section 3.3 which would ideally be combined into a one-step-process.

Third, as suggested in Section 3.2, one could find ways to use qualitative data as well. Transforming qualitative data to quantitative variables might open new opportunities in predicting syndicate success or in applying filters for excluding certain syndicates.

In general, one could think about a more sophisticated combination of ranking and filtering methods in order to come up with a list of target syndicates for participation.

Another area where it would be desirable to make refinements is the question of capital versus capacity. The common accounting methods and the inter-syndicate comparison of profitability is somewhat limited because it does not take into account the amount of capital required to back the participation on a certain share of capacity. However, the difficulty remains the availability of data in this respect. PartnerRe's "Capital at Lloyd's" team could try to use the "Member Modeller" to come up with an estimate of the marginal capital requirement for each syndicate if it were to be added to their current portfolio. But this is neither straightforward nor perfectly relevant. It would be more interesting to have the possibility to combine any set of syndicates to a portfolio and know the corresponding capital requirement.

In conclusion, the collection of data and the refinement of syndicate segmentation approaches is an ongoing process in which there are still many challenges to be solved.

3.7 The selection of target syndicates in practice

The quantitative frameworks for syndicate segmentation discussed in this work should be seen as a sort of pre-screening for a capital provider looking for the most interesting opportunities in the Lloyd's market. They can be used to obtain a ranking of syndicates based on available data and quantitative considerations. Such a ranking should then be refined by the incorporation of all kinds of other information. Especially the experienced professionals' knowledge about the market can be used to modulate the segmentation. Furthermore, a variety of strategic considerations should be used to further filter or adjust the obtained order of preference of investment opportunities.

Also, segmenting the syndicates in order to prioritize potential investment opportunities is obviously only a one-sided approach, whereas it requires two parties to conclude a capital deal. The segmentation methods presented in this project do not take into account in any way whether or not the respective syndicates are actually looking for new capital providers.

If it were possible to determine the probability to become a capital provider to a syndicate for each syndicate, this aspect could be included in the ranking. However, this is far from being straightforward.

In practice, the preferences of a potential capital provider have to be intersected with the capital needs of the syndicates. Like for other aspects of business in the Lloyd's market, this oftentimes happens through brokers. Depending on who will get in contact with whom at what moment, different deals are envisaged and negotiated.

In general, participating in the Lloyd's market is very much a people business. Many decisions involve a least some degree of human judgment which goes beyond quantitative considerations. Trust and confidence are important for the relationship between a MA and capital providers.

However, the requirement for a good connection at the human level does not mean that there would not be rigorous checks that could be performed before seizing an investment opportunity. A due diligence checklist can guide the process of evaluating different characteristics of a syndicate. Carefully looking at reserves of the syndicate or inquiring about the reputation of a syndicate in the market are examples of such due diligence activities. Similarly, for a capital provider it is important to verify that the management of the syndicate has "skin in the game", i.e. has interests that are aligned with those of the investors.

In conclusion, the selection of syndicates that should be approached for potential participation is a multifaceted challenge which requires a lot of experience and knowledge about the market.

3.8 Diversification benefit and portfolio management

Considering how capital requirements are set at Lloyd's (see Section 2.1.6), it would be interesting to maximize diversification benefit in a portfolio.

The total capital requirement for supporting the portfolio of a Member depends on how the different syndicates complement or superpose each other in the risks they are writing. This depends on both geographical and line of business diversification and is reflected for example in a comparison of the RDS percentages of the syndicates.

Lloyd's capital setting process is based on internal modeling and a set of rules as described in Section 2.1.6. It cannot be expressed in general terms by how much the required FAL of a Member will increase when this Member adds a participation on a certain syndicate to its portfolio. In some special cases it could even happen that the FAL requirement decreases.

The influence of a hypothetical participation on the total amount of capital has to be evaluated on a case by case basis with the "Member Modeller". As mentioned in Section 3.6.3 above, it would be useful to include some element related to required marginal capital into the syndicate segmentation process.

The topic of diversification benefit is further complicated by the different ways in which MAs want the capital providers to participate. There are syndicates on which a corporate TPC provider participates through its own CM, but there are also other syndicates which prefer that the corporate TPC providers take a share of the aligned, dedicated CM of the syndicate. Given that the diversification benefit is calculated at the level of each CM, this makes a significant difference.

Portfolio management is another highly important aspect for investors. As it is beyond the scope of this project, we only briefly discuss a few points here.

In general, when it comes to creating a balanced portfolio, it is important to respect at least the following three conditions:

— The portfolio should be sufficiently large.

— Its elements should be of similar size.

— Its elements should be independent of each other.

In the Lloyd's market, the size of the portfolio is certainly a major factor for well-balanced results. In practice, the number of syndicates in the portfolio of a corporate TPC provider can range from 1 to maybe 30. There are numerous aligned syndicates in the Lloyd's market, so only a fraction of the 96 syndicates are available. Also, a higher number of syndicates means many more interactions with different parties, which goes along with corresponding requirements for the number of staff managing a portfolio. For these reasons, the size of PartnerRe's Lloyd's portfolio, which is in the process of growing, might find its optimal size at about 15 syndicates.

Concerning the total amount of capacity taken on each syndicate (the size of one element), there are, as discussed in Section 3.6.1, no strict rules. Sometimes it can make more sense to look for a fixed amount on each syndicate, but sometimes the unequal size of the syndicates and available capacity results in unequal amounts of participation.

Independence of the different elements is not easy to fulfill in practice, especially in the catastrophe line of business. That is why it is important to achieve a diversified mix of business. By participating on syndicates writing several lines of business, the portfolio can be kept in balance even if one or two syndicates are quite specialized. What also matters for independence is the geographical dimension of a portfolio. The split of business by geography is therefore part of the characteristics that one should be well aware of before adding a syndicate to a portfolio.

4 Conclusion

For the conclusion of this work, we will first highlight the contributions made to the activities of PartnerRe's "Capital at Lloyd's" team, then critically discuss some limitations of the newly developed methodologies, and finally give an outlook on the future business of capital providers in the Lloyd's market.

4.1 Value created

The challenging task of syndicate segmentation of PartnerRe's "Capital at Lloyd's" team has been supported in different ways through the projects related to this *Mémoire*.

First of all, the integration of data from different sources facilitates the use of data for various purposes. One of these purposes is the quantitative syndicate segmentation as a pre-screening for interesting opportunities for PartnerRe as a capital provider in the Lloyd's market.

One advantage of the methodologies developed in this work is that they allow for a comparison across all the syndicates at once, in contrast to the APCL scoring which is available only for a minority of syndicates. The result of the segmentation process is a ranking that can serve as a basis for approaching syndicates to discuss potential participation.

The comparison of hypothetical portfolios suggests that a portfolio built with the help of the newly developed segmentation methodologies will be at least as profitable as a portfolio selected based on existing external rankings and much better than a random selection.

Given that qualitative factors play a major role in the selection of syndicates, the segmentation is only the very first step in this process.

4.2 Critical discussion

The methods for syndicate segmentation that were developed in this project are not clearly better than existing methods. Nevertheless, the development of these methods represents an independent confirmation of the rankings established by SRL and APCL. This means that in the future it is not with blind confidence but with a refined understanding that PartnerRe's Capital at Lloyd's team can use the SRL and APCL rankings as well as the newly developed segmentation methods.

The attempt to identify factors associated with syndicate profitability through statistical methods was only partly successful. While some light has been shed on patterns in the data, the findings have not been conclusive enough to establish a new "magical recipe" for syndicate segmentation. The identified factors largely confirmed the professional intuition of the people in the Capital at Lloyd's team.

The inherent difficulty in syndicate segmentation still remains that it is not straightforward to predict the future from the past.

Moreover, in the context of PartnerRe's participation in the Lloyd's market, there are a variety of constraints that further influence the choice of syndicates. The rankings based on quantitative factors are thus subject to filtering in order to exclude certain syndicates.

Overall, a combination of methods should be used to inform decisions of PartnerRe as capital provider to syndicates. More importantly, the judgments based on the experience of the people who have been working in the Lloyd's market for many years will certainly continue to play the main role in decision-making.

4.3 Outlook

PartnerRe's "Capital at Lloyd's" team will continue to broaden its portfolio of syndicates to which it provides capital. Mutual trust in the interaction with the individual syndicates will be one of the core elements of a successful long-term commitment. For the syndicates that it is backing, the "Capital at Lloyd's" team obtains much more information and data than for the others, meaning that the tracking of the syndicate performance can and should go in much more depth than the superficial quantitative analyses during syndicate segmentation.

At the same time, it is important to keep an eye the broader context and to follow the developments in and around the Lloyd's market in general [16].

Overall, the Lloyd's market remains an attractive marketplace. There are however significant differences between the individual syndicates, so it will be important to make use of successful syndicate segmentation methodologies. In many cases, profitability relies at least in part on reserve releases which are expected to decline in the medium term.

The series of hurricanes in the third quarter of 2017 have offset the recent profitability of the market. The CAL 2017 is very likely to be closing with a loss for Lloyd's as a whole. It remains to be seen how the market conditions will be impacted by these events. However, the excess capital available in the sector means that the impact on rates might be less significant than hoped for.

Brexit is not expected to affect the Lloyd's market in any fundamental way. Following the UK's vote to leave the European Union, Lloyd's has applied to the Belgian regulatory authorities to establish an insurance subsidiary in Brus-

16. For example the SRL Monitor reports are nicely summarizing the current developments, see [10] for the September 2017 version.

sels by mid-2018, but the amount of its business affected by EU passporting arrangements is limited.

There is quite a lot of activity in the Lloyd's market in terms of mergers and acquisitions. In the period 2013-2017, some 37% of the total capacity in the market have been subject to acquisitions [10]. This means that it is important to follow the dynamics of group support.

In conclusion, the Lloyd's market is an attractive marketplace for a capital provider who is able to identify and seize suitable opportunities through a complex process which starts from a multi-criteria segmentation of syndicates, ends in successful capital deals and requires the expertise of experienced professionals all along the way.

References

[1] Society of Lloyd's: *Corporate history*. https://www.lloyds.com/lloyds/about-us/history/corporate-history [accessed 11.11.2017].

[2] Society of Lloyd's: *Stability and security*. https://www.lloyds.com/lloyds/about-us/what-do-we-insure/stability-and-security [accessed 11.11.2017].

[3] Society of Lloyd's: *Economic Capital Assessment (ECA) Guidance Manual*. https://www.lloyds.com/~/media/files/the-market/tools-and-resources/member-modeller/eca-process-guidance-manual.pdf?la=en [accessed 11.11.2017].

[4] Society of Lloyd's: *Syndicate SCR for 2018 YOA. Guidance notes – June 2017*. https://www.lloyds.com/~/media/files/the-market/operating-at-lloyds/solvency-ii/2018-guidance/scr-2018-yoa-guidance.pdf?la=en [accessed 11.11.2017].

[5] Society of Lloyd's: *Market Bulletin Ref: Y5113*. https://www.lloyds.com/~/media/files/the-market/communications/market-bulletins/2017/08/y5113.pdf [accessed 11.11.2017].

[6] Syndicate Research Limited: *About Us*. https://www.syndicateresearch.com/about/ [accessed 11.11.2017].

[7] Robert Stonehewer-Smith: *Lloyd's of London: How Lloyd's Works*. Syndicate Research Limited special report, September 2015.

[8] Robert Stonehewer-Smith: *Continuity Opinion Methodology*. Syndicate Research Limited, December 2016.

[9] Syndicate Research Limited: *Appendix 3: Lloyd's Market Statistics*. 2017.

[10] Syndicate Research Limited: *Lloyd's Monitor*. September 2017.

[11] Insurance Insider (April 2017): *Names-backed syndicates post 12.1% return for 2014 YoA*. Page 18.

[12] Argenta Private Capital Limited: *Lloyd's Syndicate Profiles - 2016*. 2016.

[13] Hampden Agencies Limited: *An Introduction to Lloyd's & Membership*. Powerpoint presentation for Members. www.hampden.co.uk/files/5014/6531/9238/Presentation_slides.pptx [accessed 11.11.2017].

[14] Hampden Agencies Limited: *Introductory Guide 1 - Capacity*. http://www.hampden.co.uk/download_file/view/696/152/ [accessed 10.11.2017].

[15] Hampden Agencies Limited: *Glossary*. http://www.hampden.co.uk/hampden-agencies/client-resources/glossary/ [accessed 10.11.2017].

[16] Hampden Agencies Limited: *Hampden Performance*. http://www.hampden.co.uk/hampden-agencies/new-clients/performance/ [accessed 10.11.2017].

[17] R Core Team (2013): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[18] Charles J. Geyer (2003): *Model Selection in R*. University of Minnesota School of Statistics Course Documentation. http://www.stat.umn.edu/geyer/5931/mle/sel.pdf [accessed 10.11.2017].

[19] Jo Hardin (2013): *Multiple Regression IV - R code. Model Building*. Pomona College Course Documentation. http://pages.pomona.edu/~jsh04747/courses/math158/multlin4.pdf [accessed 10.11.2017].

[20] Bret Larget (2003): *Cp, AIC and BIC*. University of Wisconsin-Madison Statistics Department Course Documentation. http://www.stat.wisc.edu/courses/st333-larget/aic.pdf [accessed 10.11.2017].

[21] Martin A. Lindquist: *Multiple Linear Regression* and *Variable Selection*. Columbia University Statistics Course Documentation. http://www.stat.columbia.edu/~martin/W2024/R6.pdf and http://www.stat.columbia.edu/~martin/W2024/R10.pdf respectively [accessed 10.11.2017].

[22] Iain Pardoe, Laura Simon and Derek Young: *Lesson 5: Multiple Linear Regression*. Penn State University Statistics Course Documentation. https://onlinecourses.science.psu.edu/stat501/node/283 [accessed 10.11.2017].

[23] Ingo Ruczinski: *Variable Selection*. Johns Hopkins Biostatistics Department Course Documentation. http://www.biostat.jhsph.edu/~iruczins/teaching/jf/ch10.pdf [accessed 10.11.2017].

List of Figures

List of Tables

Appendix I (non-confidential)

Excerpts of code for R

Data was imported into R from csv-files in which the headers of the columns had the same names as the candidate predictors presented in Table 4. The column with the dependent variable was labeled "Profitability".

Approach based on automatic functions

The code below shows how predefined R functions were used to obtain correlation matrices and to build linear models. For those functions, `cor(...)` and `step(...)`, the data array has to be free of any blanks, so all the rows containing blanks were removed from the imported data, yielding the dataframe `theDataWithoutBlanks`.

This code was run 1) with the unimputed data and 2) with the imputed data. The corresponding outputs (from the `summary(...)` functions) are displayed in Appendix II (see page **??**).

```
# generating and exporting a correlation matrix
correlationMatrix=cor(theDataWithoutBlanks, method="pearson")
write.table(correlationMatrix, paste(labelExperiment, "
    _CorrelationMatrix.csv", sep=""), row.names=TRUE, sep=";",
     dec=".")

# predefining models with none and with all predictors (will
    be used as boundaries in step function)
nullModel=lm(Profitability~1, data=theDataWithoutBlanks)
fullModel=lm(Profitability~BasisMarket+Capacity+LogCapacity+
    GroupSupport+Alignmen+TPC+Age+LTOS+InvestmentReturn+
    ReinsEff1+ReinsEff2+Acquisitio+Admin+Forex+
    PremiumRetention+WrittenVsEarned+NLR+GLR+PercentagePaid,
    data=theDataWithoutBlanks)

# necessary for BIC
numberOfObservations=nrow(theDataWithoutBlanks)

# a) Forward selection with AIC
outputFwd=step(nullModel, scope=list(lower=nullModel, upper=
    fullModel), direction="forward")

# b) Stepwise forward with AIC
outputStepFwd=step(nullModel, scope=list(lower=nullModel,
    upper=fullModel), direction="both")

# c) Stepwise backward with AIC
```

```
outputStepBckwd=step(fullModel, direction="both")

# d) Backward elimination with BIC
outputBckwd=step(fullModel, direction="backward", k=log(
    numberOfObservations))

# printing the results
summary(outputFwd)
summary(outputStepFwd)
summary(outputStepBckwd)
summary(outputBckwd)
```

Approach based on dynamic dataset

The code below shows how a **forward selection** algorithm was implemented to avoid using the predefined R function `step(...)` in order to avoid deleting the rows with blanks a priori. The dataset is thus dynamic in the sense that the `lm(...)` function each time only ignores those rows which contain blanks *in any of the columns of the candidate predictors involved in that particular model*. The data is present in the dataframe `theData`.

Another advantage of this method compared to the predefined `step(...)` function is the generation of a customized table that tracks the adjusted R-squared, the AIC and the BIC values along the way of the model building algorithm.

```
# preparing table for overview of results
tableOfModels=data.frame(row.names = c("formula", "
    coefficient", "p-value","significance", "min p", "R2adj",
    "max R2adj", "AIC", "min AIC", "nObs", "BIC", "min BIC"))

# defining variables & setting them to right start values
y="Profitability"
jmax=19
alpha=0.05
predetPartOfFormula="" # predetermined part of the linear
    model formula, initially not present
nPriorModels=0 # the cumulated number of models that have
    been checked in earlier loops

# preparing vectors
headersx=colnames(theData[,-1])
listOfSummaries=c()
listOfCoeff=c()
```

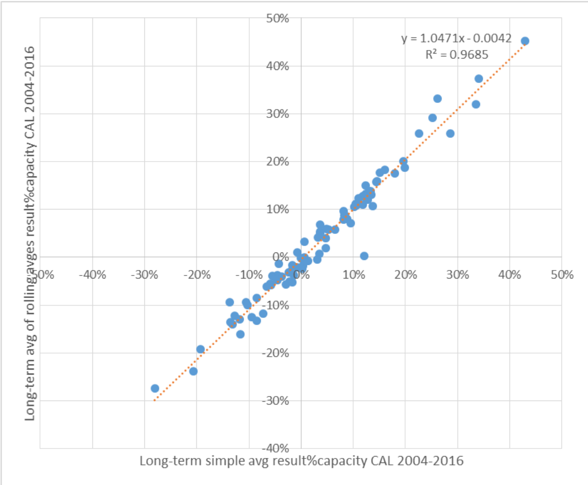
```
indexPmin=c(NULL)
indexR2max=c(NULL)
indexAICmin=c(NULL)
indexBICmin=c(NULL)

# START OF THE TWO FOR-LOOPS

# the outer for-loop which allows adding more and more
    variables to the model (forward selection)
for(j in 1:jmax){

  # the inner for-loop which goes through all the independent
      variables that are not in the model yet
  for (i in 1:length(headersx)){

    currentIndexOutput=nPriorModels+i # index within the
        tableOfModels in which the results are successively
        accumulated

    currentFormula=as.formula(paste(y,"~",
        predetPartOfFormula, headersx[i])) # in the format as
        required by lm()
    currentFormulaAsText=paste(currentFormula[2],
        currentFormula[1], currentFormula[3]) # in the format
        as required to write into the tableOfModels
    tableOfModels[1,currentIndexOutput]=currentFormulaAsText

    # collecting summary and coefficients of the linear model
        currently under consideration
    outputlm=lm(currentFormula,data=theData)
    listOfSummaries[[currentIndexOutput]]=summary(outputlm)
    listOfCoeff[[currentIndexOutput]]=summary(outputlm)
        $coefficients


    # adding certain outputs to the corresponding columns in
        the tableOfModels:

    # coefficient of the currently added variable
    tableOfModels[2,currentIndexOutput]= (listOfCoeff[[
        currentIndexOutput]][j+1,1])

    # p-value of to the currently added variable
    tableOfModels[3,currentIndexOutput]= (listOfCoeff[[
        currentIndexOutput]][j+1,4])

    # R-squared adjusted
```

```
  tableOfModels[6,currentIndexOutput]=summary(outputlm)$adj
      .r.squared

  # AIC
  tableOfModels[8,currentIndexOutput]=AIC(outputlm, k=2)

  # BIC
  tableOfModels[10,currentIndexOutput]=nobs(outputlm)
  tableOfModels[11,currentIndexOutput]=AIC(outputlm, k=log(
      nobs(outputlm)))


  # significance levels for p-value displayed in the 4th
      row/column
  if(as.numeric(tableOfModels[3,currentIndexOutput])
      <0.0001){tableOfModels[4,currentIndexOutput]="****"}
  else if(as.numeric(tableOfModels[3,currentIndexOutput])
      <0.001){tableOfModels[4,currentIndexOutput]="***"}
  else if(as.numeric(tableOfModels[3,currentIndexOutput])
      <0.01){tableOfModels[4,currentIndexOutput]="**"}
  else if(as.numeric(tableOfModels[3,currentIndexOutput])
      <0.05){tableOfModels[4,currentIndexOutput]="*"}
  else if(as.numeric(tableOfModels[3,currentIndexOutput])
      <0.1){tableOfModels[4,currentIndexOutput]="."}
  else{tableOfModels[4,currentIndexOutput]=""}

  if(as.numeric(tableOfModels[3,currentIndexOutput])<alpha)
      {tableOfModels[4,currentIndexOutput]=paste(
      tableOfModels[4,currentIndexOutput], " significant"}

}


# finding the min/max value for each criterion and printing
     a label in the corresponding row and column

indexPmin[j]=which.min(as.numeric(tableOfModels[3,(
    currentIndexOutput-length(headersx)+1):
    currentIndexOutput]))
tableOfModels[5,(currentIndexOutput-length(headersx)+
    indexPmin[j])]=paste("min p for j = ",j)

indexR2max[j]=which.max(as.numeric(tableOfModels[6,(
    currentIndexOutput-length(headersx)+1):
    currentIndexOutput]))
tableOfModels[7,(currentIndexOutput-length(headersx)+
    indexR2max[j])]=paste("max R2adj for j = ",j)
```

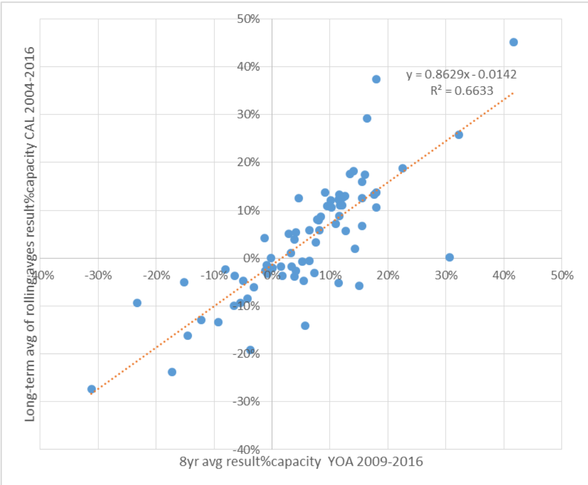
```
    indexAICmin[j]=which.min(as.numeric(tableOfModels[8,(
        currentIndexOutput-length(headersx)+1):
        currentIndexOutput]))
    tableOfModels[9,(currentIndexOutput-length(headersx)+
        indexAICmin[j])]=paste("min AIC for j = ",j)

    indexBICmin[j]=which.min(as.numeric(tableOfModels[11,(
        currentIndexOutput-length(headersx)+1):
        currentIndexOutput]))
    tableOfModels[12,(currentIndexOutput-length(headersx)+
        indexBICmin[j])]=paste("min BIC for j = ",j)


    # exiting the for-loop on j if there is no new p-value
        below the level of significance
    if(as.numeric(tableOfModels[3,(currentIndexOutput-length(
        headersx)+indexPmin[j])])>alpha){
      print("no significant variable anymore")
      break
    }

    # updating the counter of the number of evaluated models
    nPriorModels=nPriorModels+length(headersx)


    # CORE OF THE FORWARD SELECTION
    # including the most significant variable into the linear
        model before starting the next loop
    predetPartOfFormula=paste(predetPartOfFormula, headersx[
        indexPmin[j]],"+")
    # excluding that variable from the variables that will be
        checked in the next loop
    headersx=headersx[-indexPmin[j]]

  }

# END OF THE TWO FOR-LOOPS

# export of results
write.csv2(t(tableOfModels), csvExportFilename, row.names=
    TRUE)
```

Plots

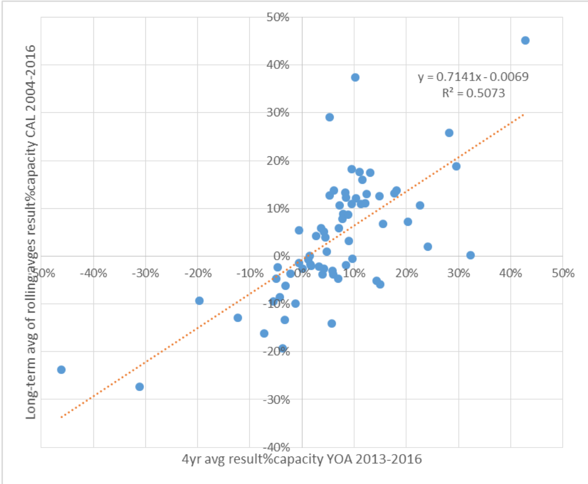
Comparing various measures of profitability



(a) Long-term average of rolling 5-CAL vs. long-term simple CAL



(b) Long-term average of rolling 5-CAL vs. 8-YOA



(c) Long-term average of rolling 5-CAL vs. 4-YOA

Figure 17 – Comparison of various measures of profitability. The long-term average of rolling 5-CAL averages is taken as the response variable. It is closely correlated to the long-term simple CAL average, less closely to the 8-YOA average and even less closely to the 4-YOA average.CAL = Calendar Year, YOA = Year Of Account

Two-dimensional plots of Profitability versus candidate predictors

For the purpose of plotting, certain data points were removed as follows:

— RITC syndicate 5678 was removed from the data for the years 2009, 2010 and 2014 due to an excessively high profitability >1

— Syndicate 1880 was removed for the year 2011 due to its loss <-1

— Syndicate 0218 was removed for the year 2010 due to its loss <-1

— For certain candidate predictors, outliers were removed from the plots. Values above $Q75 + H$ or below $Q25 - H$ with $H = 1.5 \cdot (Q75 - 25)$ and $Q25$ and $Q75$ referring to the first and third quartile respectively was removed for the candidate predictors InvestmentReturn and following in the Table 4.

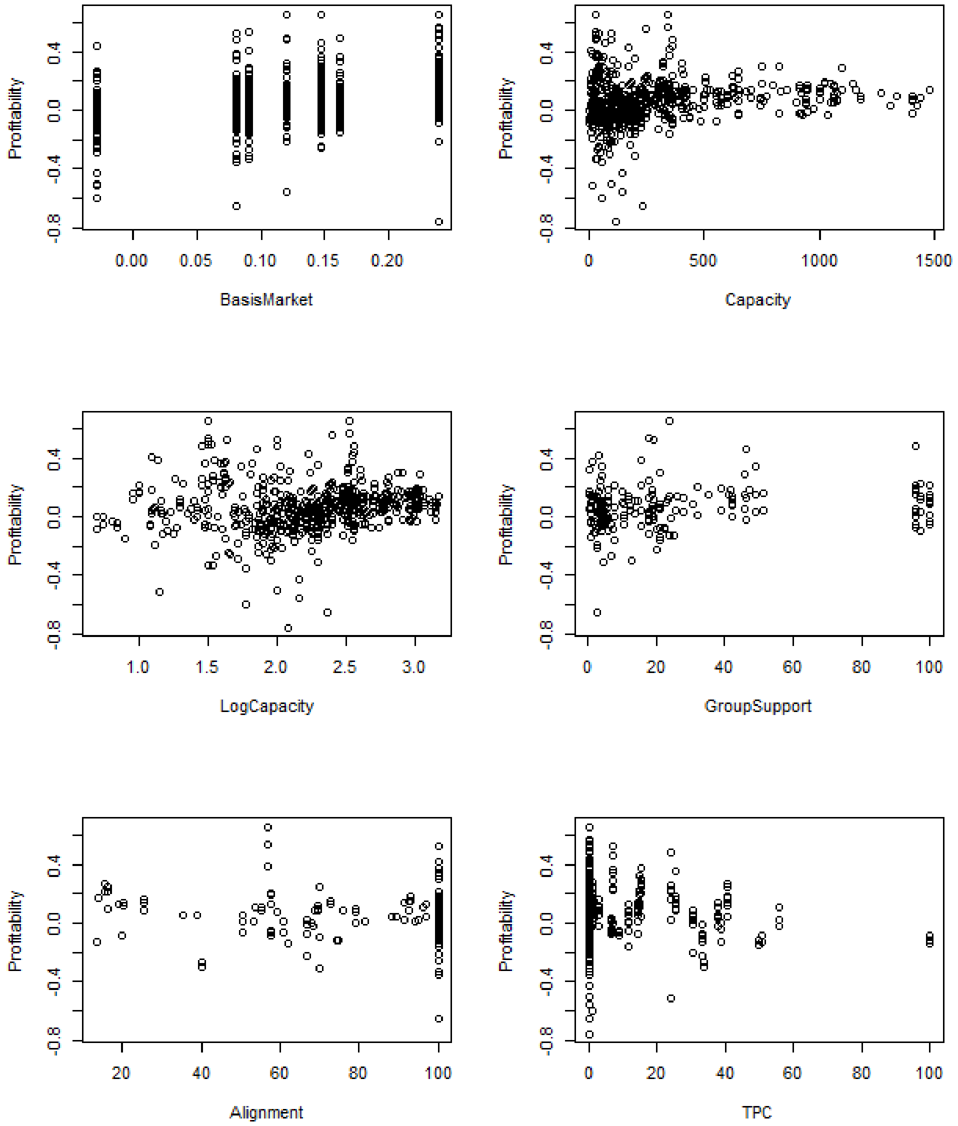


Figure 18 – Plots of Profitability versus candidate predictors for the original dataset (Part 1).

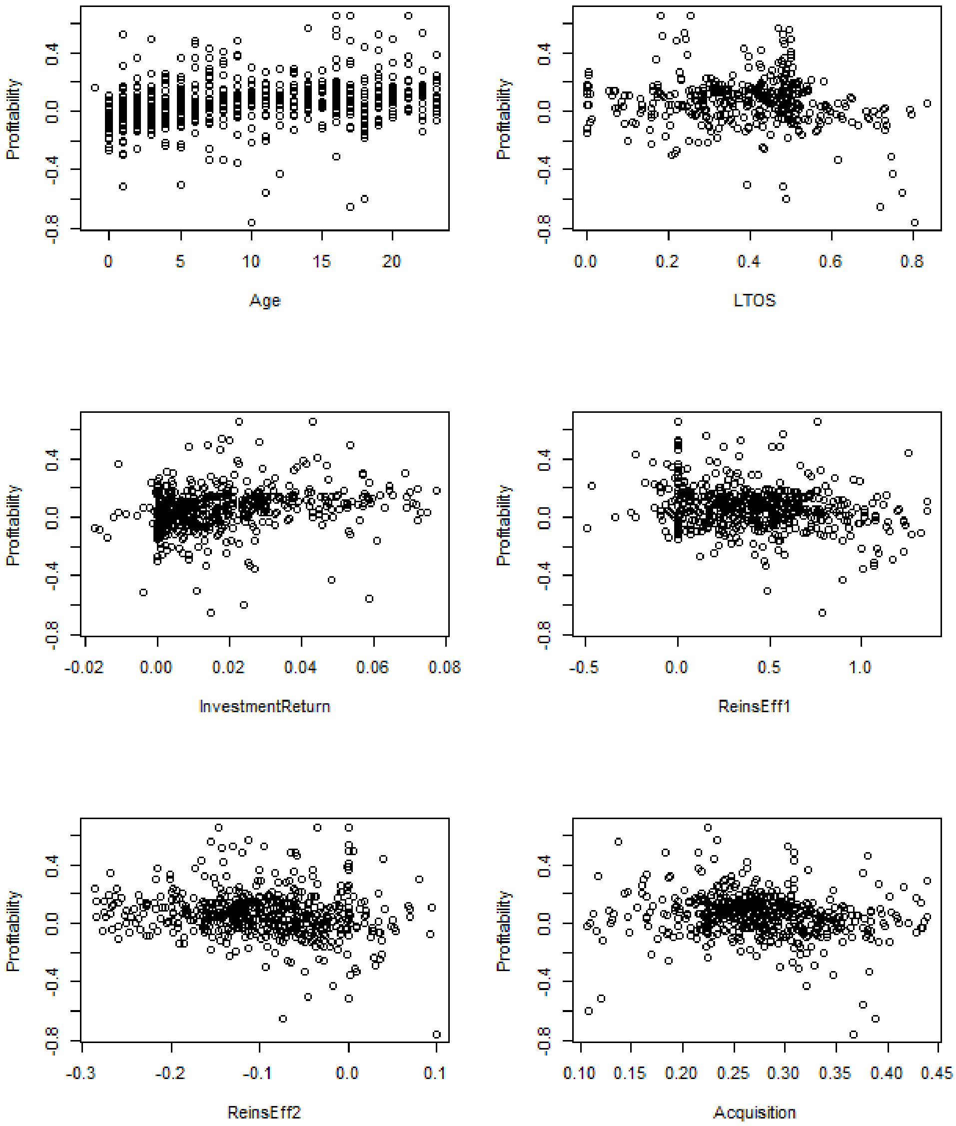


Figure 19 – Plots of Profitability versus candidate predictors for the original dataset (Part 2).

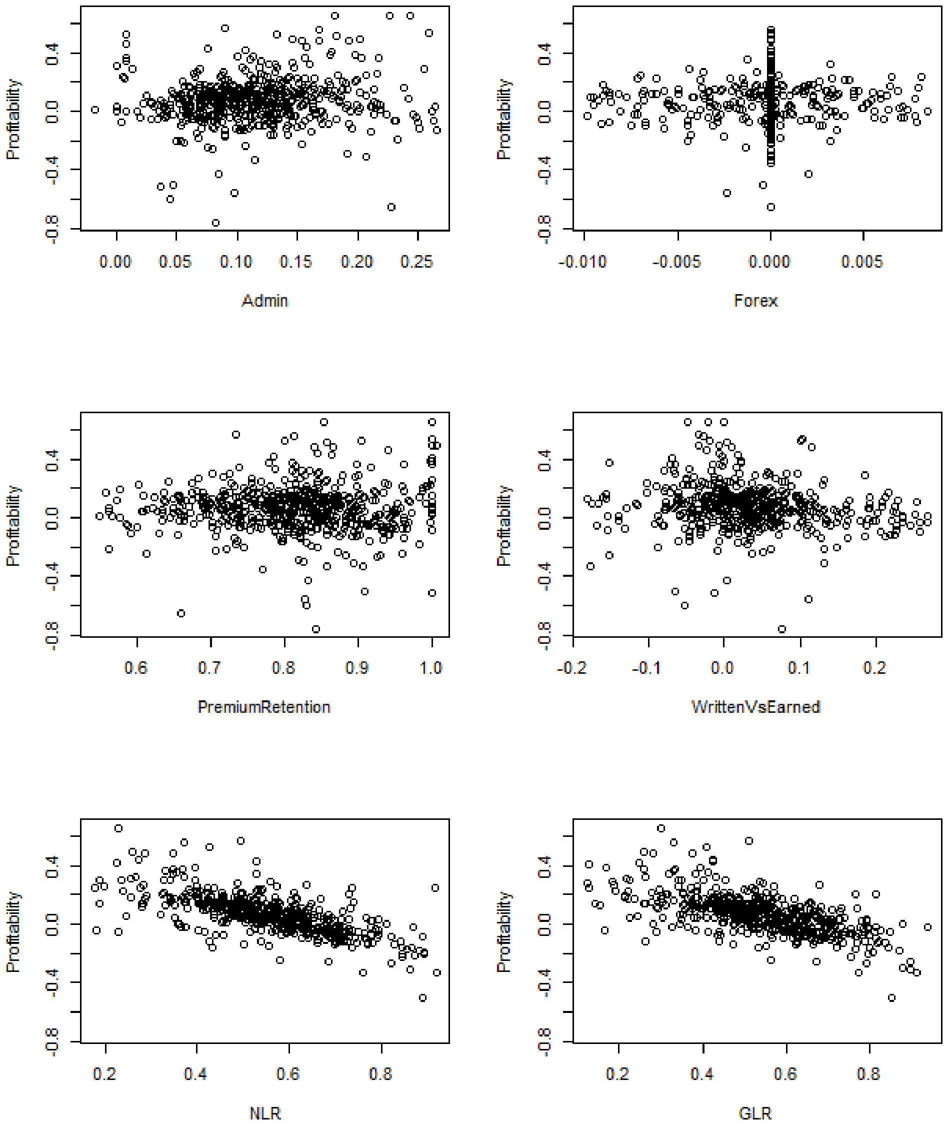


Figure 20 – Plots of Profitability versus candidate predictors for the original dataset (Part 3).