University of Southern Queensland, Australia
Faculty of Business

A Decision Model for Real Estate Portfolio Valuation and Optimisation

Under consideration of real estate physical characteristics

A Dissertation submitted by

Marvin C. Simoni, Dipl. Ing. FH, MBA

For the award of

Doctor of Business Administration

June 2010

ABSTRACT

In today's business environment, the asset evaluation models used to reach an optimised asset management situation are one of the important tools that can help a company to gain a competitive advantage. A firm's balance sheet contains different types of assets and this study focuses the analysis on the tangible and fixed asset of "real estate" (RE), which includes buildings and land.

This study is an applied research project on the topic of real estate portfolio (REP) management. It uses a cross-sectional design with the aim of developing a REP empirical decision model (REP-EDM) for a pension fund (PF) to utilise as part of its REP evaluation processes. The REP-EDM is based on the benchmarking of REP physical characteristics to a REP benchmark. Correlational research methodology with a multivariate regression is used to develop the REP-EDM model. The model is limited to the Canton Zurich in Switzerland but the methodology may be applied to other RE markets.

The relevant theories that have been considered are: Real estate theory, finance theory with the focus on investments, risks and modern portfolio theory, as well as benchmarking theory. In the literature, REP optimisation models are focused on the risk/return ratio, benefits and occupancy costs. There is limited evidence of REP optimisation models that start from an empirical model based on a REP benchmark. Thus, this research addresses a relevant topic of interest within the community that has not yet been empirically investigated. The research question has been formulated as

follows: *How can a customer's REP be optimised in order to reduce its idiosyncratic risks, basing the analysis on the REP's physical characteristics and comparing it to a benchmark of the RE market physical characteristics?*

The issue of estimating RE liquidity risk is crucial in developing a successful REP strategy and the REP-EDM including the REP benchmark contributes to extending the existing body of knowledge regarding REP management, transparency and understanding of the RE market. In the model for REP evaluation developed in this study, the interpretation of the statistical significance of the most relevant variables included into REP-EDM is done with a practical significance analysis, which includes two practical applications.

The REP-EDM can be used as an additional decision support system for PF managers in order to answer the research question of this study in an objective way and independently from RE specialists. The REP-EDM model does not substitute other REP optimisation models but instead, it represents an additional model that supports managers in taking strategic decisions in a RE market characterised by low transparency and inefficiency.

CERTIFICATION OF DISSERTATION

I certify that the ideas, experimental work, results, analyses, software, and conclusions reported in this dissertation are entirely my own effort, except where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

Signature Marvin C. Simoni, Candidate

04.06.2010
Date

ENDORSEMENT

Signature Dr Peter Phillips, Supervisor USQ

Date

Signature Prof. Dr Giampiero Beroggi, Supervisor HWZ,
Adjunct Professor University of Zurich

04.06.2010
Date

ACKNOWLEDGEMENTS

I would like to thank my academic principal supervisor Dr. Peter Phillips for his support, feedback, guidance and insight throughout this research project.

Special thanks to my academic associate supervisor Prof. Dr. Giampiero Beroggi, who shared gladly and without hesitation his valuable know-how, experience and wisdom.

Many thanks to Gregor Sasek, my coach, mentor and employer, for allowing me the opportunity to develop my expertise in real estate management and giving me the possibility to realise this study.

A lot of thanks to Katherine Dean Pirkle for her editing skills, counsel and as second lecturer during the development of the project.

I am especially grateful to my loving wife, Tina, for her continuous love, encouragement and support with our three children.

Marvin Carlo Simoni

Zürich, June 2010

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BVG	Swiss Law for Social Insurance (BVG = Berufliche Vorsorge Gesetz)
BVV2	Swiss Legal Ordinance for Social Insurance, second revision. (BVV2 = Berufliche Vorsorge Verordnung, zweite Revision)
CBD	Central Business District
CMBS	Commercial Mortgage Backed Securities
CO	Condominium (Eigentumswohnung, EWO, Stockwerkeigentum, STWE)
FZG	Swiss Law for Personal’s Own Capital (FZG = “Freizügigkeitsgesetz”)
GIP	General Investment Portfolio
GIS	Geographic Information Systems
MBS	Mortgage Backed Securities
MFD	Multi Family Dwelling (“Mehrfamilienhaus”, MFH)
MPT	Modern Portfolio Theory, also referred to as MVA
MRE	Multiple Regression Estimation
MSAs	Metropolitan Statistical Areas
MVA	Mean Variance Analysis, also referred to as MPT
PF	Pension Fund (“Pensionskasse”)
PMPT	Post Modern Portfolio Theory
RE	Real Estate
REIT	Real Estate Investment Trust
REP	Real Estate Portfolio
REP_S	REP of a single pension fund
REP_B	REP Physical Characteristics Benchmark of a Specific RE Markets
REP-EDM	REP Empirical Decision Model
SFD	Single Family Dwelling (“Einfamilienhaus”, EFH)
SPI	Swiss Performance Index
STA	Statistical Office of the Canton Zurich
SWX	Swiss Stock Exchange, Zurich, Switzerland
TQM	Total Quality Management
ZKB	Zurich Cantonal Bank (Zürcher Kantonalbank), 8010 Zurich, Switzerland
ZWEX	Residential property price index for the Canton Zurich (ZWEX = Zürcher Wohneigentumsindex)

LIST OF FIGURES

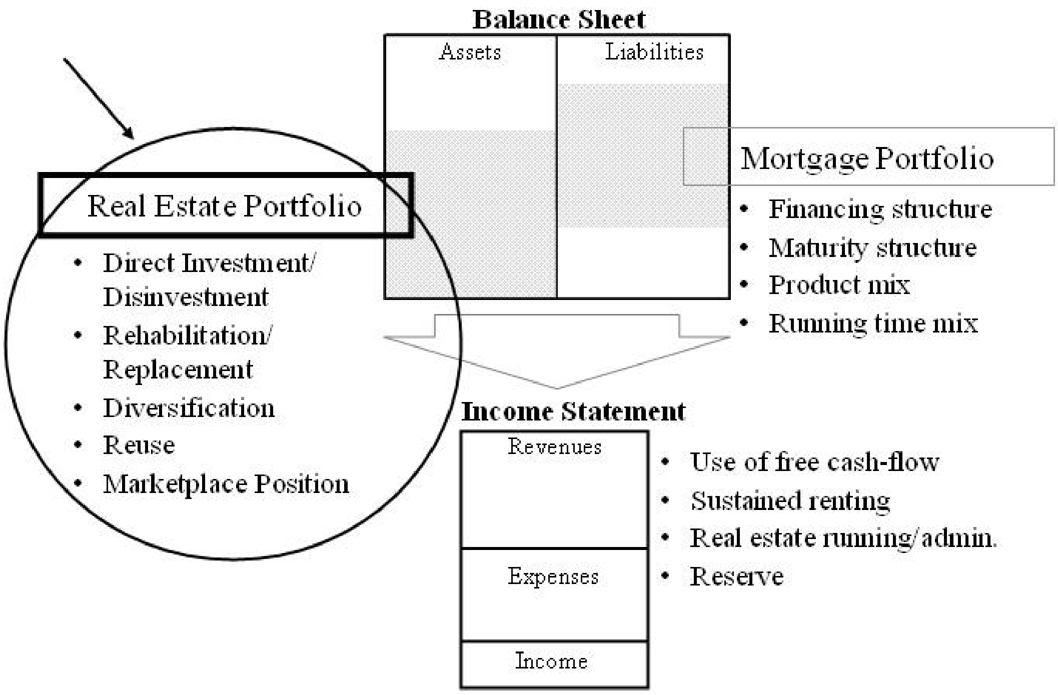
LIST OF TABLES

1 INTRODUCTION

1.1 Introduction

The asset evaluation model used to reach an optimised asset management situation is one of the important that can be employed to gain a competitive advantage. As depicted in the Figure 1.1, the asset and the liability values influence the income statement of a firm and therefore its result. A firm's balance sheet (Figure 1.1) contains different types of assets that can be classified into many different groups. One possible group contains the tangible assets (also referred as fixed assets or PPE - property, plant, equipment) and includes assets such as real estate with land and buildings, plants, furniture and machinery.

Figure 1.1 Area of Interest and Position of this Study



(Source: Developed for this research)

This study focuses on the analysis of the tangible asset "real estate" (Figure 1.1). The aim of the investigation is to develop a decision model for real estate portfolio (REP) evaluation under consideration of real estate (RE) physical characteristics. Primary goal of the research and analysis is to develop a methodology for benchmarking a pension fund's REP against a set of RE physical characteristics, the REP_benchmark model drawn from the market. The benchmarking model developed in this study can be used both as an instrument to evaluate a customer's REP and as an instrument for decision-making in the REP optimisation process.

In the existing literature, that is, the material found through extensive research, the investigated REP evaluation models focus on the risk/return ratio, benefits and occupancy costs. The author has found no evidence of REP evaluation models that start from an empirical model based on a REP_benchmark as in this study. Thus, this investigation addresses a relevant topic of interest that has not yet been empirically investigated.

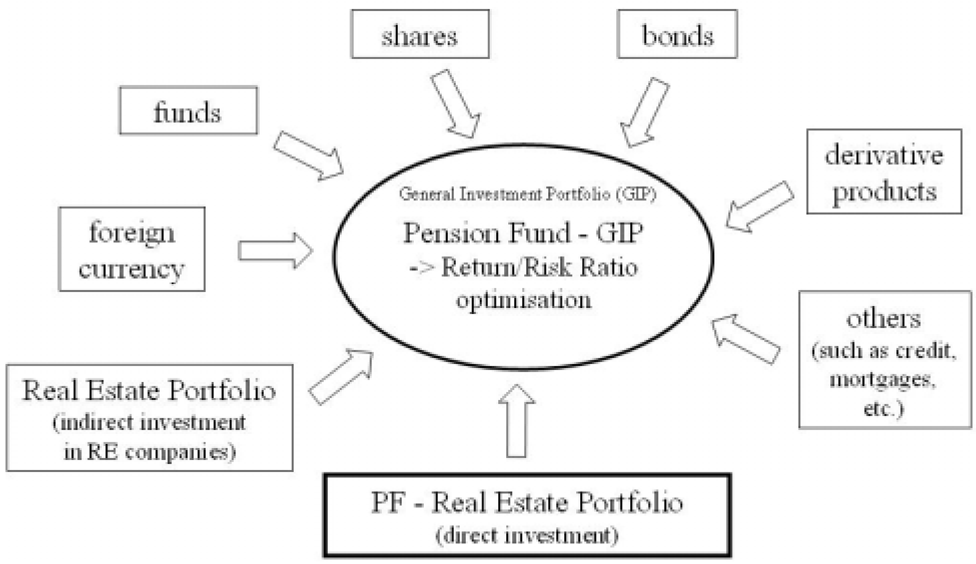
1.2 Overview of the Chapter

The purpose of this chapter is to introduce this dissertation. In Section 1.3, the background of the research is presented. In Sections 1.4, 1.5 and 1.6, the research problem, contribution, justification and planned methodology are highlighted. In Section 1.7, the outline of the report is documented while controversial terms are defined in the Section 1.8. In the Section 1.9, the delimitations of the study are presented and in Section 1.10, the key points are summarised in a conclusion.

1.3 Background to the Research

A general investment portfolio (GIP) may contain shares, bonds, funds, credit such as refinancing products, mortgages, derivative products, direct real estate (RE) investments such as building and land, as well as indirect RE investments in the form of participation and/or equity in RE companies (Figure 1.2). In the RE area, different models exist which can be used to optimise the risk/return ratio in connection with a general investment portfolio (GIP). In those GIP optimisation models, a real estate portfolio (REP) is often analysed as a diversification factor to optimise financial aspects and to reduce the related investment risks. Using these models it is possible to optimise the return, adapting the weights of the mentioned assets in the portfolio, depending on the risk aversion of the customer (Sing & Ong 2000; Fama & French 1996).

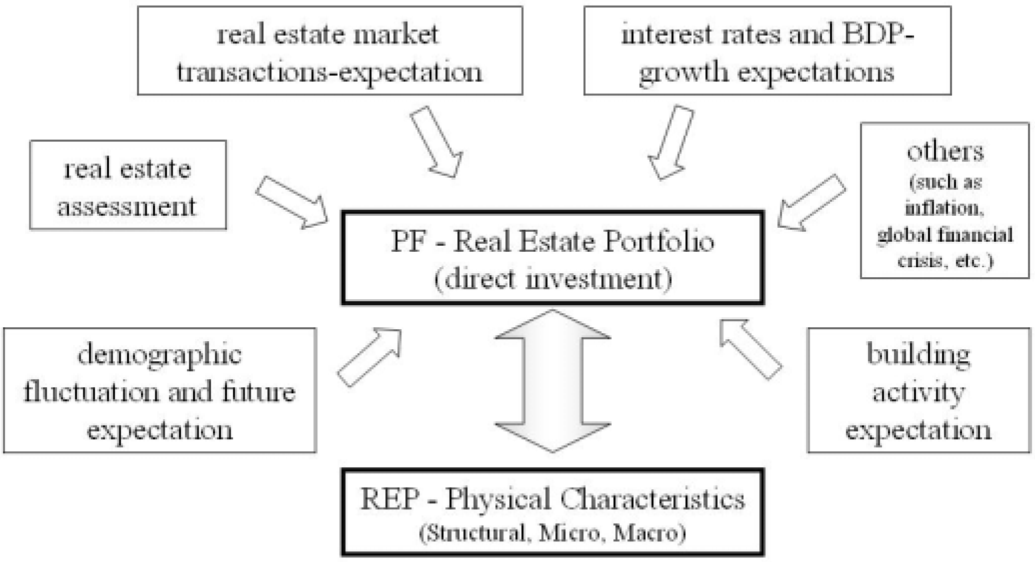
Figure 1.2 General Investment Portfolio (GIP)



(Source: Developed for this research)

In such models, the REP (direct investment) is often considered as a single object with all its internal parameters which include its physical characteristics and its RE-specific factors such as RE market transactions-expectation, RE assessment, inflation, economic influences like the global financial crisis on the RE market, etc. (Figure 1.3). REP optimisation models, in turn, attempt to optimise the risk-return ratio, basing the analysis exactly on the internal parameters of the REP itself (Benjamin et al. 2001). This study is focused on "REP Physical Characteristics" (Figure 1.3). A REP_benchmark is developed against a set of RE characteristics. The REP_benchmark may be used by practitioners as the basis for REP decision-making.

Figure 1.3 Real Estate Portfolio (REP) - Physical Characteristics



(Source: Developed for this research)

Customer questions such as "How well does my RE portfolio reflect the physical characteristics of the REP_benchmark (market)?" and "How can I optimise my RE portfolio to reduce the risks compared to the benchmark of the RE market?" can be answered today by RE specialists who base their knowledge only on their experience with local RE markets. This fact has been acknowledged by the CEO of the Pension Funds of the Zurich Cantonal Bank (PF ZKB), Raymonde Hiltmann, in an interview

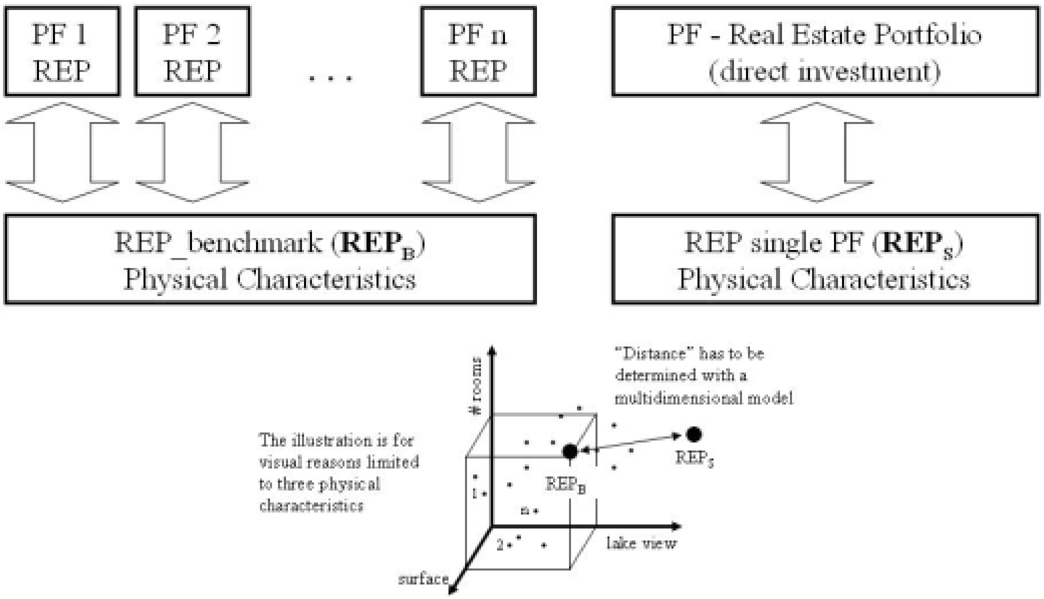
(Hiltmann 2007). She confirmed that the REP assets of the PF ZKB were valued at the end of 2006 at about 327 million CHF, which included 48 multiple dwellings, and that this portfolio was managed by RE specialists that based their decisions only on their personal knowledge of the RE market.

Geltner and Miller (2001, p. 534) confirm '…the need for specialised local expertise when investing in property assets.' Montezuma (2004) notes that managing residential assets requires property specialists with knowledge in a wide range of RE related issues. Schulte et al. (2005, p. 95) add that '…local knowledge and information about the local market is essential for the successful purchase or development of property.' In searching for answers to the above mentioned questions, there appears to be no empirical or theoretical model that considers the relationship between the physical characteristics of the customers' RE portfolio and a benchmark of the RE market.

The purpose of this study is to develop an objective REP empirical decision model (REP-EDM) for the evaluation of a customer's REP under consideration of the REP physical characteristics such as lake view, surface, number of rooms, age, location. The main goal is to develop and test a methodology for benchmarking a REP to a set of RE physical characteristics. Thus, the principal focus is not to find out the causes and reasons of the current RE market situation, but rather to evaluate a customer's REP compared to a REP_benchmark (REP_B). The "distance" between a customer's REP_S and the benchmark of the physical characteristics of the RE market (REP_B) can be considered as a measure of the idiosyncratic risks of the customer's REP_S. The greater the "distance" to the benchmark, the greater the predicted idiosyncratic risks for that investment (Figure 1.4).

The risk carried by a pension fund (PF) is defined in this study as a function of the PF's REP "distance" to the REP_benchmark (Figure 1.4), where the REP_B stands for the benchmark and the REP_S is a single PF's REP. According to this definition, if a firm has its REP_S close to the REP_benchmark it will have a higher possibility of completing a transaction. This higher possibility of completing a transaction is due to higher trading-related liquidity, and thus a smaller risk of not being able to sell or buy the object. This, however, does not imply that companies should be risk avoiders.

Figure 1.4 REP_benchmark based on RE Physical Characteristics



(Source: Developed for this research)

According to the portfolio theory, a rational investor should avoid the idiosyncratic risks that are not priced in the market because such are able to be diversified away. This "distance" is a measure that may be used for the evaluation of a REP_S and sets the baseline for the customer's decisions for an optimisation of the investments in the REP_S with the possibility to replicate the RE market, reducing the idiosyncratic risks of the customer's REP_S.

1.4 Research Problem

Given the background to the research, the following research problem is advanced:

How can a customer's REP be optimised in order to reduce its idiosyncratic risks, basing the analysis on its physical characteristics (REP_S*) and comparing it to a benchmark of the RE market physical characteristics (*REP_B*)?*

1.5 Justification for the Research and Contribution

Geltner & Miller (2001, p. xxiv) affirm that the '…commercial property market is not as liquid or efficient in its operation as the securities markets.' Support for this position is made by the following authors: Braun et al. 2008, Topintzi et al. 2008, Montezuma 2004 and Amman & Scherer 2001. Possible reasons for this lower liquidity of the RE market compared to the securities markets are because RE assets cannot be bought and sold quickly and the transaction costs are not irrelevant (Georgiev et al. 2003). Despite the low liquidity, in Switzerland the RE market reflects the biggest asset class with approx. 2.5 trillion CHF at the end of 2006 (Maier et al. 2008). As a comparison, the capitalisation of the Swiss Performance Index (SPI) for the same period reached 1.48 trillion CHF (SZC 2008; Halbherr 2007). Therefore, it is important for customers and managers to know where and how they should invest money into their REP and these decisions should not be based solely on the RE specialists' experience but should be based also on an objective REP empirical decision model (REP-EDM) because every real estate decision has long-term consequences (Apgar 1995).

Apgar (1995) affirms that what managers need is a process that they can use to diagnose whether their REP has a competitive position as well as a set of tools to facilitate their leadership role in real estate decisions that have to be linked with the business strategy. He says that managers recognise that by managing RE as a business function, they can cut costs significantly, increasing at the same time productivity, and that before a manager can decide what he wants and where he wants to go, he has to know what he has and where he is. This situation could be supported with a REP-EDM that could help managers in the analysis of their current RE situation. As mentioned by Hiltmann (2007), the development of such a model based on a REP_benchmark would increase market transparency in the RE business facilitating and supporting strategic decisions.

A REP-EDM that includes the physical characteristics of a REP is relevant for analysing the position of a customer's REP_S in relation to the benchmark on the RE market (REP_B). In fact, when a customer has a defined REP target, this model permits an evaluation of its REP_S and allows the customer to optimise, to increase or decrease the "distance" of his REP_S to the REP_benchmark, as desired to close the gap between the existing investment portfolio REP_S and the strategic REP target. Another contribution of this study is to provide a model which will contribute to enhancing the understanding and the transparency of the RE market, building a REP_benchmark into a defined area and providing the possibility to compare it with a specific REP_S. This will permit the calculation of the risk level (as measured by the "distance" between the REP_S and the REP_B) of the analysed REP_S compared to the benchmark of the RE market (REP_B) and will generate an objective answer to the research question.

The motivation for the research derives from the potential utility of the resulting REP evaluation model. The model will support decision-making by determining a REP physical structure that is relevant for the strategy chosen. By following the risk-averse strategy of RE market (REP_B) replication it is possible to generate a reduction of REP risks that emerge as a result of divergences from the RE benchmark. Less risk averse strategies may be based on decreasing replication of the benchmark. Such a model will contribute to the body of knowledge of RE management and will increase the quality of consultants' services, which in turn will contribute to winning new customers and strengthening customer ties through additional REP analysis.

1.6 Methodology

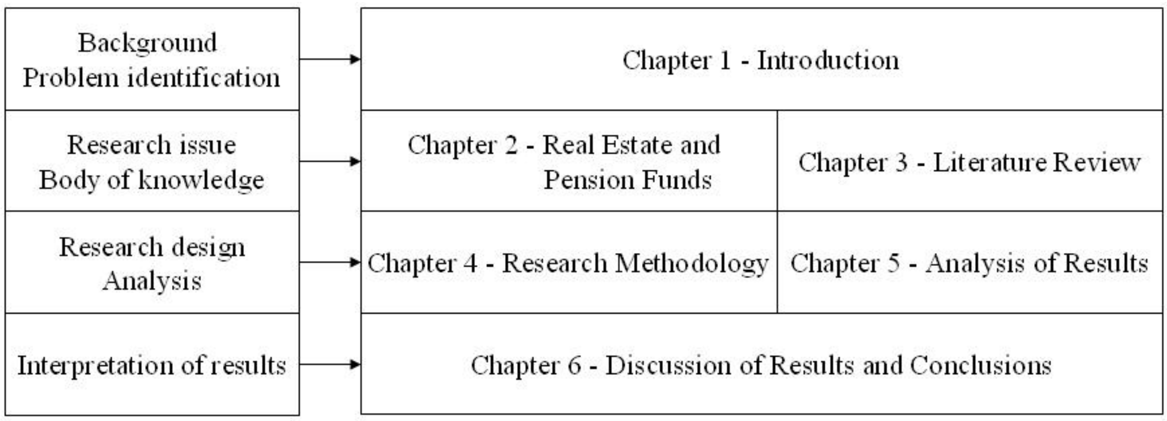
This study is an applied empirical research project on real estate portfolio (REP) management and uses a cross-sectional design to obtain quantitative data with the aim to develop a REP empirical decision model (REP-EDM) for a customer's REP evaluation under consideration of its RE physical characteristics and the REP_benchmark. The principal focus of this study is not to discover the causes of and reasons for the current RE market situation, but rather to evaluate a PF's REP compared to a REP_benchmark (REP_B), basing the analysis on existing data. Firstly, the comparison of REP_S itself justifies the use of the correlational research method for this study. Secondly, no control group can be used with or created from historical data and, thirdly, the involvement of metric data promotes the use of this methodology.

The procedure used for this study, described with complete details in the Chapter Four, contains various steps starting from the data collection and consolidation, going through to the description of the important physical factors (variables) for REP characterisation, the computation of the REP_benchmark, the definition of the "distance" as measure between two RE portfolios and the development of the REP-EDM. The quantitative analysis of the data is conducted using SPSS (Statistical Package for Social Sciences) and additional statistical tests considered as scientifically appropriate are used to ensure validity and reliability of the developed model.

1.7 Outline of the Report

The structure of this report is based on the format used in quantitative research proposed by Perry (2002) and on the three central questions (what, how and why) mentioned by Punch (2006), which a research needs to answer. A six-chapter structure has been developed to present this dissertation. The six-chapter structure is depicted in the following Figure 1.5.

Figure 1.5 Outline of Dissertation



(Source: Developed for this research)

Chapter One, Introduction, outlines the broad field of the study and leads into the focus of the research problem with its background. It gives an overview of the methodology, it includes a definition of controversial terms and it provides a delimitation of scope.

Chapter Two, Real Estate and Pension Funds, presents the context in which the study takes place including the Zurich pension funds management industry, the prominence of RE and diverse figures and statistics on the size and trends of the RE industry.

Chapter Three, Literature Review, builds the necessary theoretical foundation by reviewing the field of the research problem and concentrates on benchmarking, specifically on investment portfolios and RE portfolios.

Chapter Four, Research Methodology, discusses and justifies the research design and the methodology used to collect the field data to address the identified research issues.

Chapter Five, Analysis of Results, presents results derived from the data analyses.

Chapter Six, Discussion of Results and Conclusions, presents findings for the research issue within the context of prior research examined in the literature review. The last section of this chapter exhibits implications and suggestions for further research.

1.8 Definitions

In this section the most important terms that are used repeatedly in the chapters to come are defined in order to establish positions taken in the research.

1.8.1 Asset

An accepted definition of asset is the one used by the International Accounting Standards Board (IASB) which affirms that '…an asset is a resource controlled by the enterprise as a result of past events and from which future economic benefits are expected to flow to the enterprise' (IASB 2005). In this dissertation, the term "asset" is used to refer only to real estate as a fixed asset unless otherwise noted.

1.8.2 Investor

In this dissertation, the term "investor" is a synonym for an economic agent who makes rational asset allocation decisions on the basis of the information revealed by asset prices or, what amounts to the same thing, exchanges assets on RE markets. A rational investor will accept a higher risk only when this is reflected in the expected profit in form of an additional reward (Hurni & Stocker 1996).

1.8.3 Real Estate (RE)

According to Geltner and Miller (2001), the investor would answer the question, "What is real estate?" by saying that, "A real estate is potential future cash flows." In this dissertation, the term "real estate" is used to refer only to a specific object that includes a building and the land where it is located. In other words, real estate encompasses land along with anything permanently affixed to the land, such as buildings, specifically property that is stationary, or fixed in location.

1.8.4 Direct and Indirect RE Investment

In a direct RE investment, the RE units, also called unsecuritised properties, are traded directly in the private property markets. In an indirect RE investment, the RE units, also called securitised properties, are traded indirectly in the public stock markets through the equity shares of real estate investment trusts (REITs) and other RE firms or investing in the secondary mortgage market such as in commercial mortgage-backed securities (CMBS).

1.8.5 Liquidity

Crouhy et al. (2001) divide the liquidity risk into funding liquidity risk and trading-related liquidity risk. In this dissertation, the term "liquidity" is used to refer only to the trading-related liquidity, which in this case is the capability to convert a real estate into cash within a short time period. This can be done selling the RE on the market.

1.8.6 Benchmarking and Benchmark

Benchmarking can be defined as a *continuous improvement process* during which processes and methods of operational functions as well as products and services of one's own company are measured against a benchmark, i.e. the maximum achievable performance (Falk 2000). A *REP_benchmark* in this dissertation is used as a *point of reference* for a measurement, therefore it can be seen as a standard against which something can be measured or assessed.

1.8.7 Distance and Idiosyncratic Risk

Zikmund (2003, p. 564) defines the '... beta as the appropriate measure of the systematic risk' of a portfolio. The market has a beta of one and the beta of individual portfolios is a measure of the risk of the portfolio *vis-à-vis* the market. Similarly, the developed REP_benchmark for the RE market, allows the determination of the idiosyncratic risk of a REP i.e. the trading-related liquidity risk of a REP portfolio, by permitting the measurement of the "distance" of the REP to the REP_benchmark. In this respect, distance to the REP_benchmark is defined, for the purposes of this investigation, as a measure of idiosyncratic risk of a REP.

1.8.8 Surface

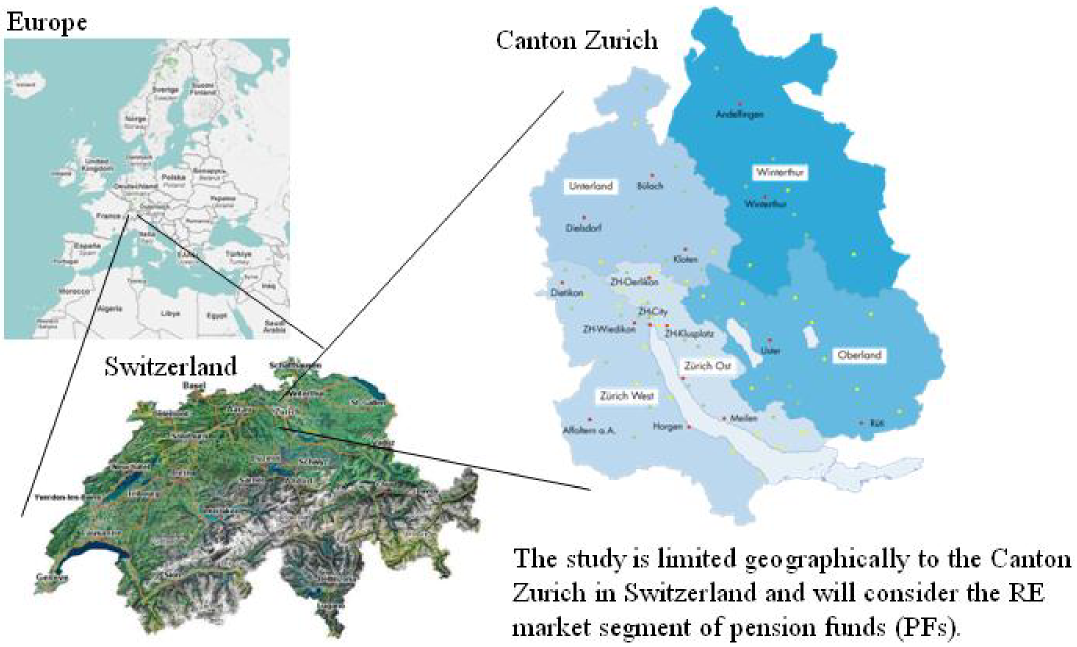
In this dissertation, the term "surface" corresponds to the usable surface in square metres for the building and more precisely it includes the fraction of the net floor area for the intended use of the building, i.e. net floor area reduced by circulation areas (corridors, stairs etc.) and functional areas (WCs, storage rooms, etc.) unless otherwise noted.

1.9 Delimitation of Scope and Key Assumptions

This study has a number of limitations embedded in its nature and scope. The REP-EDM, including the REP_benchmark built in this study, is delimited geographically to the Canton Zurich in Switzerland and considers the RE market segment of pension funds (PFs) with domicile in the Canton Zurich. In Switzerland, there are approximately 2,200 registered PFs (SFG 2008d) and, in the Canton Zurich alone, there are approximately 750 (Fuhrer 2006). This delimitation, due to data protection, is given by the availability of the RE raw data of the RE market segments to be analysed. In fact, due to confidentiality of data it is difficult to obtain any relevant information for the research. Nevertheless the researcher's professional involvement with statistical office of the Canton Zurich (STA) provides him with access to relevant data for PFs of the Canton Zurich only (Figure 1.6).

In brief, this study involves PFs in a specified region, thus the results reflect the RE market situation in that region and on a specific date as defined by the cross-sectional design. This geographical limitation, reduces the interpretation of the results only on the Canton Zurich, in other words the REP "distance" measurement of risk can be applied only for this canton. Therefore the results cannot be generalised to apply in a larger context. However, the methodology developed within this dissertation may be applied to other markets and, of course, this study provides an additional decision supporting system that can be used by PFs willing to invest in Canton Zurich but not resident in it.

Figure 1.6 Geographical Limitation of the Study



(Source: Developed for this research)

In this study, it is assumed that the raw data received from the statistical office of the Canton Zurich (STA), the building assurance Zurich (GVZ), the Swiss federal government for statistics (SFG) and the Zurich cantonal bank (ZKB) has the necessary quality to guarantee validity and reliability. Thus, for example the reliability of the data collection methods used by the various sources is assumed to be given.

1.10 Conclusion

Chapter One, Introduction, has established the foundations for this dissertation. In the first chapter, the scene was defined and set and the path along which the reader will travel towards the thesis' conclusion was outlined. It introduced the research problem with its background, presented the justification for the research, gave an overview of the methodology, included the key definitions used in the dissertation and finally, it provided a delimitation of scope. On these foundations, the report can proceed with a detailed description of the study. In Chapter Two the real estate and the pension fund markets are presented.

2 REAL ESTATE AND PENSION FUNDS

2.1 Introduction

Chapter One identified the research question and objective. This chapter introduces the environment in which this study takes place by demonstrating the importance of the real estate (RE) asset class in the business world, showing the role of the pension funds in the direct RE investments, indicating the specialities of the RE vis-a-vis other goods, pointing out the two different and possible RE direct or indirect investments, and presenting a discussion of the segmentation of the RE market. The presentation of the study's context starts with an overview of the real estate (RE) market in Switzerland and the Canton Zurich and proceeds through the RE characteristics and the pension funds (PFs) including various figures and statistics on the size and trends of the RE and PFs industry. The discussion of the context is drawn to a close with a summary and conclusion.

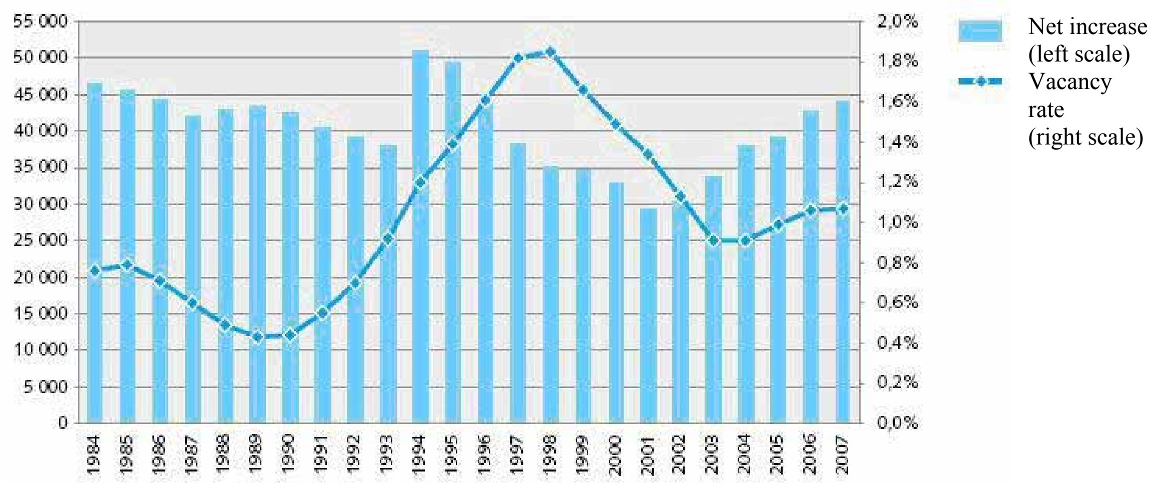
2.2 Overview of RE Market in Switzerland and in Canton Zurich

Switzerland, with about 7.5 million residents, lies in the centre of Western Europe. With a surface area of 41,284 square kilometres, it belongs to the smaller nations in Europe. The north-south expansion amounts to maximally 220 kilometres, in west-eastern direction lies the maximum with approximate 350 kilometres. In 2007, of the total land area in Switzerland, 36.9 percent were agricultural areas, 30.8 percent forest and wood areas and 6.8 percent RE areas. The last 25.5 percent of the total surface area was unproductive surface such as water, unproductive vegetation and vegetation-less areas (SFG 2008b).

Although the RE area represents the smallest percentage of Switzerland's total area, at the end of 2006, the RE market comprised the biggest asset class with approx. 2.5 trillion CHF (Maier et al. 2008). As a comparison, the capitalisation of the Swiss Performance Index (SPI) for the same period reached 1.48 trillion CHF (SZC 2008; Halbherr 2007). This assets comparison is a demonstration of the importance of the RE industry in a country. The magnitude of the market means that the solvency of RE owners can influence the economic cycles of expansion, prosperity, contraction and recession of the nation.

A trend exhibited by the Swiss RE market is the increased requirements for dwellings and mobility between 1995 and 2005. This has led to a growth of the RE area to nine percent per year by 2008 (SFG 2008b) and to an increase in the number of employees in the RE sector from about 20,000 to about 30,000 (SFG 2008a). A second trend is the building of increasingly larger apartments. The mean surface area available per person increased between 1990 and 2000 from 39 to 44 square meters (SFG 2008a). A third trend concerns the supply and demand in the RE market of residential properties. A key measure on the supply side is the net increase in the number of new-built homes. This measure, which comprises newly built homes adjusted for increases and decreases caused by conversions and demolitions, measures the growth in the housing stock. A possible indirect measure on the demand side is the vacancy rate, which is a figure indicating the absorption by the market. The Figure 2.1 indicates that the increased supply in 2006 and 2007 has been compensated by an increased demand, in fact the vacancy rate remained almost constant (SFG 2008c).

Figure 2.1 RE Market for Residential Properties in Switzerland



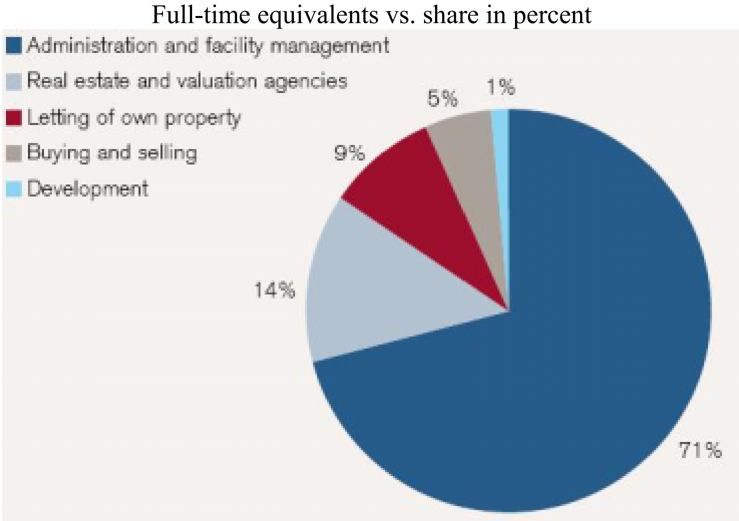
(Source: SFG 2008c)

Switzerland's RE market is divided into a host of submarkets and the regional RE markets differ from one another. Socio-economic mismatches are a characteristic of each region. They have roots in various driving forces such as the differences in the location factors such as accessibility and resources, the deregulation of the financial markets and the liberalisation of goods and service markets. Thus, aggregated national figures such as the RE surface percentage or such as the vacancy rate depicted in the Figure 2.1 do not match with all regions in the same manner. The result manifests itself in very different regional and segment-specific values (SFG 2008c).

The next issue to be considered is the variety of services available relative to RE. The official statistics divide the RE industry into five subsectors as showed in the Figure 2.2. The subsector administration and facility management includes sell and buying activities and is the dominant subsector with 71 percent of all employees in the RE industry. Real estate brokerage and valuation agencies is the second largest subsector with 14 percent of employees followed by the subsectors of companies that let out their own properties, firms whose main business is to buy and sell their own properties and finally the smallest subsector, that is mainly concerned with developing land (Braun et al. 2008).

Figure 2.2 Employment by Real Estate Services

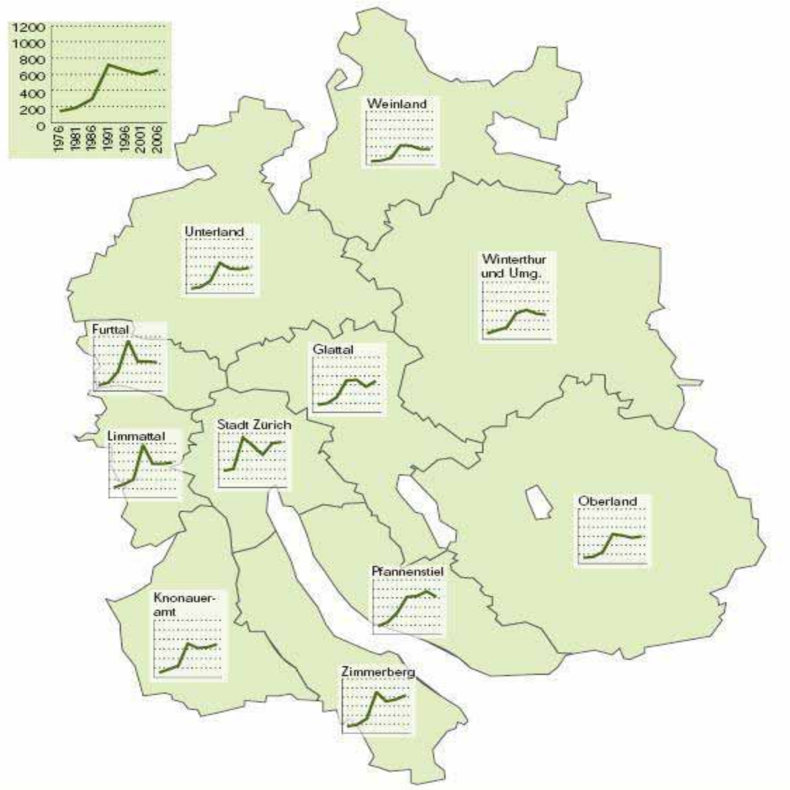
Full-time equivalents vs. share in percent



(Source: Braun et al. 2008)

The Canton Zurich, with about 1.3 million inhabitants and a surface of 172,871 hectares, is the most densely populated canton of Switzerland. In 2007, the RE surface in this canton reached the 20.1 percent of its total area and included an RE asset value of more than 350 billion CHF. The Canton Zurich consists of 171 political municipalities and eleven regions (Figure 2.3). These regions are divided according to geographical characteristics. They combine neighbouring municipalities into areas that are as structurally uniform as possible (STA 2008a; STA 2008b).

Figure 2.3 Ground Prices Development in Canton Zurich



(Source: STA 2008b)

The Zurich RE market is characterised by a number of observed trends. The first trend is the price development of the empty ground available for building among its regions. From the Figure 2.3, it can be seen that the price development is very different from one region to another. In fact, in 2007, the price spectrum started from the Weinland region with a mean price of 350 CHF per square meter, extending to the city of Zurich region, with a mean price of 1000 CHF per square meter. In the regions Furttal and Limmattal a relative high mean price can be observed and the explanations of this price development could be the topography and the distance to the Zurich centre. For the regions Pfannenstiel and Zimmerberg an additional explanation for the relative high mean price could be the lake view (STA 2008b).

Another trend is the price development of residential properties over time, depicted in the Figure 2.4, through the residential property price index for the Canton Zurich (ZWEX). The development since 1980 can be subdivided roughly into three main phases. In the 1980s, the RE market boomed. Over ten years, RE prices in the Canton Zurich doubled. During the RE crisis of the 1990s, the index dropped approximately 50 points below the former top value. Since then, the prices for residential property have risen again, but still have not reached the past top values from the late 1980s (ZKB 2008).

Figure 2.4 Residential Property Price Index for the Canton Zurich (ZWEX)

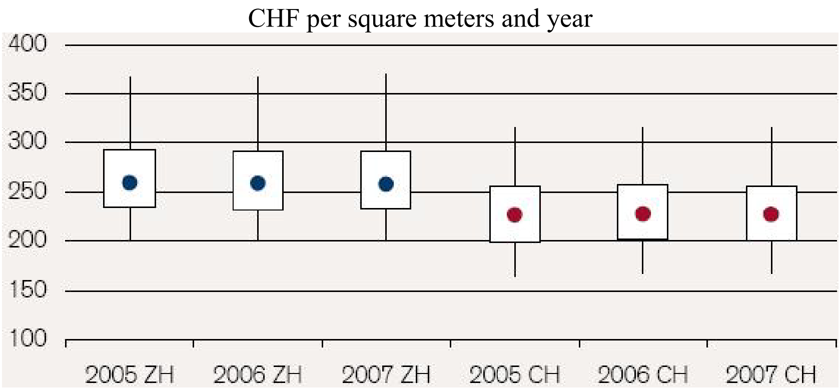
(Source: ZKB 2008)

The next observation concerns the vacancy rates in the canton. In 2007, the vacancy rate for residential properties was at 0.8 percent. This is a lower percentage than the Swiss mean of 1.07 percent, indicating a balanced RE market between supply and demand. The same cannot be said for the vacancy rate for office properties. In fact, their vacancy rate in 2007 was 4.6 percent; high compared to other similar cantons in Switzerland, such as

Geneva and Bern with 1.1 percent, Basel with 2.6 percent and Lausanne with 1.8 percent, thus indicating a possible demand problem in Zurich (Braun et al. 2008).

The third trend to be described is the development of the advertised rents for residential properties in CHF per square meter and year. The Figure 2.5 shows that the rents in the Canton Zurich (blue dot) lie over the rents in Switzerland (red dot) over time. The block shows the range between the 30th and 70th percentiles. The black line joins the 10th and 90th percentiles, and the blue/red dot corresponds to the median (Weber et al. 2008).

Figure 2.5 Development of Rents

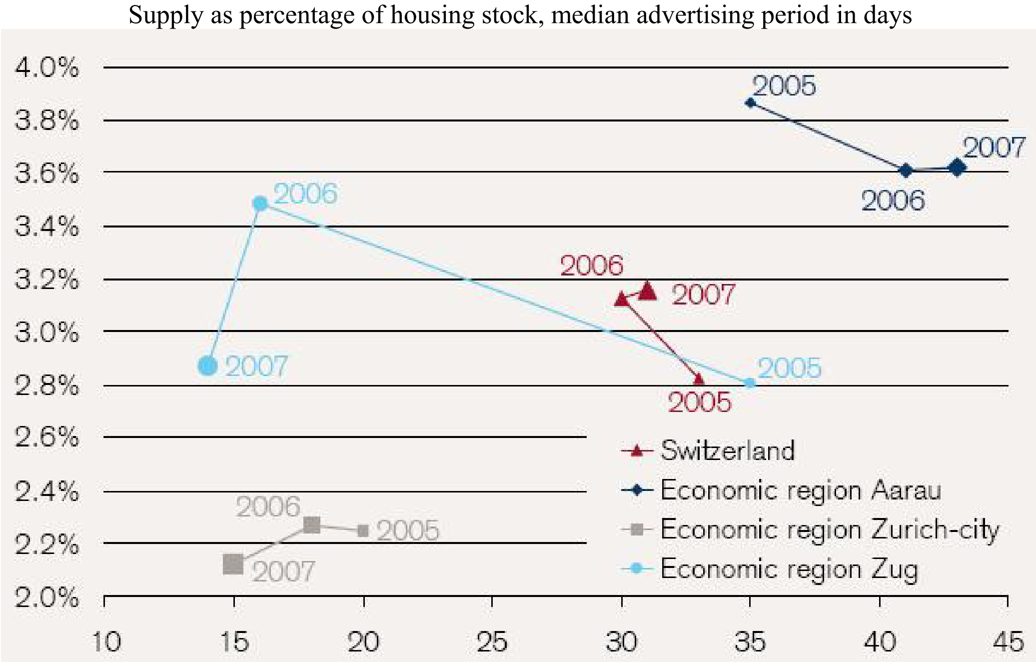


(Source: Weber et al. 2008)

The final trend to be described is the RE market liquidity as a function of supply rates and advertising periods. Braun et al. (2008, p.22) state that a measure which reflects the market's liquidity is the supply rate. The supply rate represents the percentage of RE that are on the market during a specific period. They affirm that the average advertising period (the time for which a RE is advertised before a buyer is found) supplies a measure of how well the market is absorbing supply. A long advertising period means that the market has a supply overhang and it is not functioning perfectly. Two possible explanations for this situation are that there are too many properties for sale on the market or that the properties offered do not meet the market requirements. The differing regional real estate markets conditions are also confirmed by the different regional supply rates and advertising periods as shown in the Figure 2.6. In fact, in the Aarau region, higher supply combined with longer advertising periods is a symptom that the market was not able to absorb all the new properties quickly enough. In the Zurich region, the opposite is the case. At first, the high demand reduced advertising periods because the available properties were taken off the market quickly and second, although there are more and more new buildings, the supply rate dropped, indicating high demand and a liquid market.

Figure 2.6 Regional Supply Rates and Advertising Periods

Supply as percentage of housing stock, median advertising period in days



(Source: Braun et al. 2008)

The Zug region is also undersupplied. In 2006, more properties were made available to the market, as indicated by a supply rate of 3.5 percent, but at the same time the advertising periods have shortened. In 2007, the market became even tighter because the number of new buildings dropped, reducing also the average advertising periods to just 14 days. In Switzerland, it can be noted that despite the increase supply rate, the liquidity of the RE market has improved since 2005 (Braun et al. 2008).

2.3 Real Estate Characteristics and Investment Possibilities

A main difference between real estate (RE) and other goods is the fact that RE is absolutely fixed to a single location. In fact, some of the more important parameters determining the value of a particular RE are the characteristics of its location. These can be broken down into macro-characteristics and micro-characteristics. *Macro characteristics* include the distance to urban centres, taxes, the quality of the environment, etc. *Micro characteristics* are more concerned with features determined by the RE's immediate position. They may include a lake view, the direction in which a property faces, the exposure to noise, the intensity of the sunshine, the distance to public transportation such as bus, etc.

Because it is fixed in one location, the benefits conferred by a property depend mainly on the qualities of the location where it is situated. An additional consequence to the fact that a RE cannot be moved is that regional imbalances in the market cannot be realigned as quickly as they can in the case of other, more mobile goods. For example, a high level of demand for second homes in popular tourist destinations leads to clear price increases because the supply of such homes is relatively inelastic on a short time scale (Braun et al. 2008; Hurni & Stocker 1996).

RE is also heterogeneous in nature (Georgiev et al. 2003). The *structural characteristics* of a RE such as size, number of floors, number of rooms, surface, volume, etc. combined with its macro and micro characteristics make each property unique to some degree. This lack of homogeneity in the RE market is a reason for the existence of various different submarkets, a reason for the difficulty in comparing one RE with another and for justifying the fact that properties with similar structural-characteristics can have high differences in their price depending on their location. In addition, the difficulties in making direct comparisons create major problems in pricing a RE and '…as a result the market is very illiquid, which can lead to large price fluctuations. Even professional property assessors can arrive at valuations of the same property that differ by up to 20 percent' (Braun et al. 2008, p. 7). Additional characteristics of RE are long product life cycles, long development time and complicated transaction (buying and selling) modalities. In fact, buying or selling a building tends to take several months due to the involvement of multiple parties such as buyers, sellers, brokers, legal representation, government offices, banks and so forth, whose delays, services, fees and commissions contribute to high transaction costs. In brief, the mentioned reasons justify why an excess of supply on the RE market cannot be cleared quickly and imbalances on the market can last for a long time.

An investor in RE has, in principle, two possibilities to allocate his money. The first is the *direct RE investment* (direct purchase) and this could include its management over time. The second is the *indirect RE investment* that contains at least the following four alternatives: a) The acquisition of shares of RE holding companies such as real estate investment trust (REIT), makes it possible to invest in RE with a relatively liquid instrument, b) the acquisition on the market of portions of one professionally managed RE fund, c) the investment in the secondary mortgage market such as in commercial mortgage-backed securities (CMBS), and d) the buying of derivative products such as indexes on RE (Ammann & Scherer 2001; Georgiev et al. 2003; Montezuma 2004).

For the RE investor, diversification is a key element when trying to optimise RE investment risk/return. With direct RE investments, diversification breaks down into two main components: geography and type of usage. Even in an age of globalisation, RE markets are highly regional. Two identical houses can stand only a few kilometres apart and yet be subject to a very different interplay of supply and demand and, consequently, have very different values or prices (Braun et al. 2008; Amman & Scherer 2001). All of these factors are important areas of consideration for those fund managers who invest in RE. Among these fund managers, managers of pension funds (PFs) figure prominently.

2.4 Pension Funds

A pension fund contains a pool of assets that are bought with the contributions to a pension plan for the exclusive purpose of financing pension plan benefits. It is established by a corporation, union, government entity in order to facilitate and organise the investment of employees' retirement funds, contributed by both the employer and employees. The pension fund is a common asset pool meant to generate stable growth over the long-term, and provide pensions for employees when they reach the end of their working years and commence retirement.

Pension funds (PFs) take the form, according to the Swiss Law for Social Insurance

"BVG art. 48 sec. 2", of either a special purpose entity with legal personality (such as a trust, foundation or corporate entity) or a legally separated fund without legal personality, managed by a dedicated provider (PF management company) or other financial institution on behalf of the fund members. At the end of 2006, there were approximately 2200 registered PFs in Switzerland with about 3 million insured persons and an entire direct RE investment of about 76 billion CHF (SFG 2008d). In the Canton Zurich alone, there were approximately 750 registered PFs (Fuhrer 2006).

The pension funds (PFs) are regulated by Swiss law and can take different forms. Independent of their forms, all Swiss PFs that offer the minimum social security according to the Swiss Law for Social Insurance (BVG) have to be registered with the Swiss federal government office for social insurance and must have statutes (Table 2.1). The PFs with statutes i.e. PF type1 to type3 are subject to the Swiss Legal Ordinance for Social Insurance, second revision (BVV2, art. 53) that defines the rules for their investments including direct RE investments (Helbling & Leutwyler 2007).

Table 2.1 Pension Funds and Legal Obligation

PF services and duties \ PF Types	PF type1	PF type2	PF type3	PF type4
Minimum social security - offered	yes	yes	no	no
Additional social security - offered	yes	no	yes	no
Registration (BVG art.48 sec.1)	required	required	not required	not required
Statute and Swiss law - FZG	**required**	**required**	**required**	not required
Swiss law - BVG	required	required	required for investment	not required
Swiss law - BVV2 art. 53	**required**	**required**	**required** for investment	not required

(Source: Developed for this research)

Pension funds control relatively large amounts of capital and represent one of the largest institutional investors in many nations. Some pension funds employ their own fund managers; others delegate responsibility to external fund managers. Invariably, they will try to achieve a diversified portfolio of investments, some in low risk areas and others in high risk areas. Actuaries determine how much is going to have to be paid out to pension holders in forthcoming years and the pension fund has to try to achieve a rate of return on its capital that will meet, or better still exceed, this target.

In order to reach the defined target, at least two factors influence PF investment decisions. The first factor is the long-term investment period available to reduce the effect of the volatility, respectively the risk of the investments, and thus permit the acquisition of a higher portion of products such as shares, which in the short-term would imply a higher risk. The second factor is the liquidity that a PF has to guarantee in order to fulfil the designated payout of benefits. The PF liquidity requirements depend on the age structure of the PF's members. A higher average age of its members means a shortened investment period because payments to retired employees must be made in the short-term. The higher the average age of the members, the higher the risk aversion of the PF. In summary, the return generated by a pension fund is dependent on the asset allocation, investment style and selected products, all of which depend on the pension fund's risk capacity and risk tolerance. Thus, a performance comparison among PF should be done with a risk-adjusted analysis in order to be able to compare the results (Hurni & Stocker 1996).

2.5 Conclusion

Switzerland's RE market is a segmented market and each region has its idiosyncrasies. The main reason is because RE is a heterogeneous good that is tied to a fixed location. Another reason is the relatively inelastic supply that does not permit an increase the RE market supply in the short run. RE market analysis can be done at different levels, starting from an international or national level going down to the canton level, the municipality level and the RE level. This study focuses on direct RE investment and analyses the RE market for registered pension funds (PFs) through the RE characteristics. Thus it includes in the analysis the various levels starting at the canton level, going down to the RE level.

3 LITERATURE REVIEW

3.1 Introduction

Chapter Two presented the context in which this study takes place. This chapter, Literature Review, builds the necessary theoretical foundation by reviewing the literature relevant to the research problem. The literature review is concentrated on benchmarking and investment RE portfolios. In this study, the role of theory is to set the basic knowledge in order to be able to answer the research question and to develop the REP-EDM. This takes place under the following theoretical framework.

- Real estate theory specifically for urban economics (space market), financial economics (capital market, asset market) and real estate system.
- Finance theory specifically for portfolio optimisation theory and theory of risk (risk/return ratio, credit risk, liquidity risk, operational risk, market risk).
- Benchmark theory specifically for REP_benchmark.

Through the literature review, this theoretical framework is investigated in depth with the aim to confirm the significance of this study and to gain more knowledge that can be used to solve the research problem.

3.2 Real Estate: Background

The following four subsections are built with a bottom-up approach. The Subsection 3.2.1 starts with the economic view of RE; the Subsection 3.2.2 indicates the unique particularities of a RE; the Subsection 3.2.3 presents the three market models in the RE system; and the Subsection 3.2.4 presents a discussion of the four categories of capital asset markets and the importance of the direct RE investments as an asset class for portfolio managers.

3.2.1 Economic View of Real Estate

In order to begin the analysis of the RE system, the following question should be answered: "What is real estate?" According to Geltner and Miller (2001), the answer depends on the profession of the person that has been asked. In fact, as an example, an architect would likely describe real estate from an aesthetic and functional perspective, an engineer would likely describe it from a physical structural perspective and a lawyer would likely describe it as a bundle of rights and duties associated with real property. These different answers imply that real estate can be studied from several different perspectives. For the scope of this study, the discipline used is that of economics.

Within the economic study of RE, there are two major branches: a) Urban economics which involves the study of cities, including the spatial and social phenomena relevant to understanding the RE and b) Financial economics which studies capital markets and the financial services industry. These two branches of economics are the most relevant for understanding commercial property from an investment perspective and, according to

Geltner and Miller (2001), the investor would answer the question "what is real estate?" by saying that a real estate is potential future cash flows. The nature of these cash flows, their magnitude, time and risk will fundamentally be determined in the rental market or space market, which is where urban economics comes in. On the other side, the potential stream of future cash flows generated by real estate can be seen as a capital asset that trades on the capital and asset markets, which is where financial economics comes in.

3.2.2 Unique Particularities of a RE

A further question in the analysis of the RE system could be: "Which are the features that make RE unique and different from both the typical corporate finance and the mainstream investments field?" In answer to this question, there are numerous factors.

<u>Liquidity</u>

- One difference according to Okunev et al. (2002, p. 182) could be that the '…RE market is not as liquid, as efficient and may be sluggish in prices when compared to the stock market' and this statement is also supported by: Braun et al. (2008), Topintzi et al. (2008), Montezuma (2004), Georgiev et al. (2003), Geltner and Miller (2001) and Amman and Scherer (2001).
- The next difference is the simultaneous existence of two parallel asset markets in which RE trades. Real estate units are traded directly in the private property markets and they are traded indirectly in the public (stock) markets through the equity shares of REITs and other RE firms. In fact, the physical (direct) property market provides the underlying asset base for the securitised (indirect) property market (Geltner & Miller 2001; Montezuma 2004).

- An additional difference is that transaction costs in buying and selling property are much greater than those in the securities market (Georgiev et al. 2003).

Localisation, geography, demographics

- Geltner and Miller (2001) add another aspect of RE, saying that because both supply and demand are location and type specific, RE space markets are highly segmented and therefore rental prices for physically similar space can differ widely from one location to another. In RE, the primary geographic units of space market segmentation are areas also known as metropolitan statistical areas (MSAs) that encompass a central city with its surrounding suburbs also called central business district (CBD) and tend to be relatively integrated economically and socially.

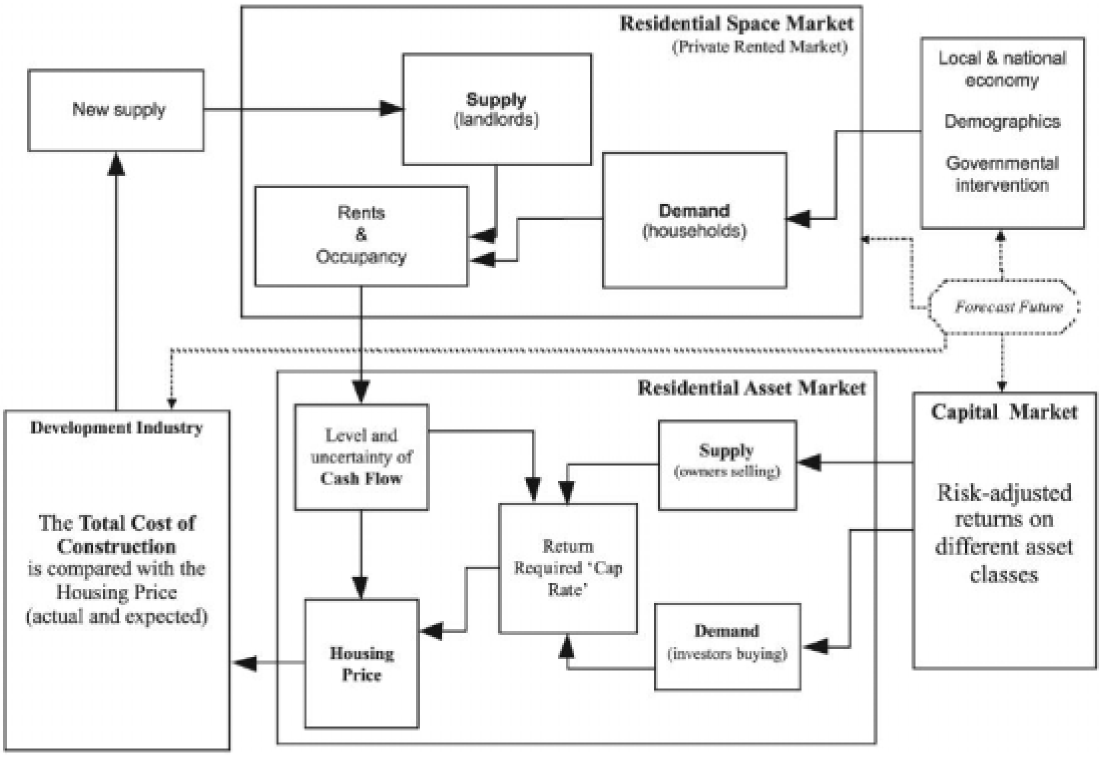
Usage

- In addition to geographical segmentation, RE space markets are segmented by property usage type. The major types include office, retail, industrial and single family and multifamily residential.
- The next speciality of the RE built space is its extreme longevity. Indeed, compared to most other products, buildings last much longer, thus rarely is a building torn down within less than 20 or 30 years from the time it is built and after all, land underlies all RE and land exists practically forever.

3.2.3 Three Market Models in the RE System

The "three markets model" is used by Archer and Ling (1997), Geltner and Miller (2001) and Montezuma (2004) to describe the linkages between the three markets for space, assets and capital. Figure 3.1 presents a visual overview of this model.

Figure 3.1 Real Estate System: Three Markets Model



(Source: Montezuma 2004, p. 232; adapted from Geltner & Miller 2001)

The *space market or rental market*, where the usage of physical space is traded, determines, depending on the space demand and supply level, the current rental levels (rents) and the expected cash flows of the RE. In this market, the demand comes from the households that use residential space. The demand for residential space depends on various factors such as: a) rent, b) demographics e.g. the number of households, its composition, age structure, c) regional and national economies e.g. interest rates, credit restrictions, income, etc., d) governmental intervention e.g. subsidy to the rented sector, direct provision of social housing or public renting that could be a substitute to private renting, and e) technological factors e.g. transport and communication network features that change the meaning of distances and urban location.

On the supply side of the space market are the landlords: Individuals and institutions that produce housing services using the property stocks available. The new supply depends on two major factors: a) the cost of developing new housing stock including the land cost, construction cost, and b) the developers' cost of equity capital. From an economics view point, a new RE will be added to the current RE space market stock when the RE value, defined in the asset market, is equal to or exceeds the marginal cost of new development.

In the literature, various RE space market analyses seek to quantify and forecast the supply and demand side of specific space usage markets, typically including the forecast of future rents and vacancies, in a particular geographic RE market segment and using the support of geographic information systems (GIS). The aim of these analyses is to help in making specific RE decisions, including investment decisions. Although these analyses tend to increase the RE market transparency, they do not include a comparison with a REP_benchmark based on the physical characteristics of a specific RE market.

The *asset market or property market*, where the available housing assets are allocated among competing investors, determines the RE value or price. RE supply comes from actual housing owners willingness to sell their RE stock. On the demand side of the asset market are the investors e.g. individuals and institutions such as pension funds. For the investor, RE assets represent a series of contingent cash flows over time, whose amount and timing are defined in the space market. Thus, RE assets compete in the capital markets with other assets promising a stream of future cash flows such as fixed-income securities, e.g. bonds (Montezuma 2004).

Geltner and Miller (2001) assert that the performance of RE assets in the *capital market* is determined by the perceptions of potential investors concerning the level and uncertainty of the assets' cash flows. In other words, the risk level of the expected cash flows earned by residential property is influenced not only by the degree of uncertainty about future relative supply and demand for RE space, but also by the co-variability of the expected cash flows with risk factors such as inflation and the interest rate, which are determined in the capital market. Thus, the risk premium of direct RE investment depends on the risk profile of the cash flows defined in the space market and its relationship with the capital market.

The *linkage between these three markets* can be described by stating that asset markets determine the value of property assets and this, in turn, governs the flow of financial capital produced by REs; this flow being influenced by the capital market. The space market can be viewed as an extension of capital market not only due to the fact that RE assets can be seen as an alternative portfolio choice available to investors but also because financing terms available on capital markets have a significant effect on the return on housing. The space market and asset market are linked to the direct short-run relationship that translates current property cash flow to current property asset value and to the medium-to-long-run by the commercial property development industry that converts financial capital into physical capital, thereby governing the stock of supply in the space market (Archer & Ling 1997).

According to the asset market approach, when the RE markets are in equilibrium, prices reflect the present valuation of future rents. The prices can be approximated as earnings (net rents) divided by the cap rate. The cap rate is calculated with information from the asset market, for example opportunity costs of capital, taxes, holding costs and risk premium in order to bear property risk. When rental and property markets are in disequilibrium an arbitrage is possible. In fact, if the cost of ownership is high, people will tend to rent, other things being equal. This relationship of equilibrium is shown in the Formula 3.1, which states that the rent is equal to the capital cost of owning a house (Geltner & Miller 2001).

$$Price * (i + \tau + f + \pi - g) = Rent \tag{3.1}$$

Where

Price	: value of the property
i	: capital costs, costs of forgone interest that the owner could have earned, net of the tax relief granted to home-owners in most countries
τ	: property taxes
f	: recurring holding costs, depreciation and maintenance
π	: risk premium for holding a risky asset
g	: expected capital gains on property
Rent	: amount paid by the tenant at specified intervals in return for the right to occupy or use the property of another

3.2.4 Four Categories of Capital Asset Markets

An overview of the major types of *capital asset markets* with their investment products is depicted in Table 3.1. The table shows how the capital markets can be divided into four categories according to whether they are public or private markets and whether the assets traded are equity or debt.

Table 3.1 Major Types of Capital Asset Markets

	Public Markets	**Private Markets**
Equity Assets	Stocks REITs (indirect) Mutual funds	Private firms *Real property (direct)* Oil and gas partnerships
Debt Assets	Bonds Mortgage Backed Securities (MBS) Money instruments	Bank loans Whole mortgages Venture debt

(Source: Geltner & Miller 2001, p. 13)

Public markets are those in which many buyers and sellers are generally simultaneously participating in the market with price quotes available for all to observe and are characterised by a relatively high degree of liquidity in that it is generally possible to quickly sell units of the assets at or near the last quoted price. This liquidity is both a cause and an effect of the fact that, in public markets, asset prices can adjust rapidly to relevant news about their value. This is market efficiency. Fama (1970, p. 383) states that, '…a market in which prices always fully reflect available information is called efficient.'

In contrast, *private markets* are those in which the assets are traded in private transactions arranged between individual buyers and sellers and are characterised by less liquidity than public markets. It takes longer for sellers to find buyers, it is more difficult to ascertain a fair price for a given whole asset and the transaction costs are typically higher in private asset markets. The fact that whole assets are traded in private deals between one buyer and one seller also has consequences for the nature of the asset price information that is available to the public. According to Georgiev *et al*. (2003) this low transparency of the RE marketplace results in potential asymmetric information. In other words, '…private asset markets tend not to be as informationally efficient as public markets' (Geltner & Miller 2001, p. 13).

Indirect RE investments in public markets provide investors with a liquid exposure to real estate via standardised financial securities in an organised, efficient and transparent market where frequent transaction-based data is readily available. At this point, the following question could be posed: "Why should investors make direct RE investments in a private market if the public market has so many advantages?" Georgiev et al. (2003, p. 29) answer this question noting that '…direct RE investment offers some diversification benefits to an established stock and bond portfolio, while securitised RE may not. Thus investment in shares of RE investment companies does not substitute for direct RE investment.'

Hurni and Stocker (1996) affirmed that real estate is a favourable investment for the investor due to the fact that real estate correlates relatively weakly with shares and bonds and so offers sufficient protection in relation to inflation shocks. In fact, important diversification benefits can be reached whenever the returns to various assets have a low correlation with one another. The returns generated by RE display a low correlation with returns generated by stocks or bonds. As a result, the addition of RE to a portfolio of stocks and bonds reduces the variability of returns to the overall portfolio and thus enables the investor to reduce risk without reducing return (Hurni & Stocker 1996).

Another special concern, mentioned by Geltner and Miller (2001, p. 534), is that

> ... in traditional RE investment, as distinguished from securities, there is the need for specialised local expertise when investing in property assets, both in the acquisition/disposition phases and in the property ownership/management phase. It may be difficult, especially for small investors to acquire sufficient expertise efficiently in more than one local area or property type. The segmentation of RE space markets and lack of informational efficiency in property asset markets, then exposes “novices” or “outsiders” to greater risk.

Geltner and Miller’s (2001) supposition, also confirmed by Hiltmann (2007), is another aspect that justifies the need for the development of an objective REP-EDM to help investors in making quicker investment decisions without having to be an expert in the local RE market. The characteristics such as lower liquidity, segmentation of RE space market and inefficiency in the private markets are the relevant drivers for this study which aims to increase the understanding and the transparency of RE markets by developing an objective REP-EDM built on a REP_benchmark.

3.3 Finance: Background

In the existing literature, numerous studies can be found regarding general investment portfolio (GIP) optimisation models in connection with RE investments as a diversification factor, with RE assessment or price estimation models and with REP optimisation in connection with the risk/return ratio. The same cannot be said about the relationship between REP optimisation models and their physical characteristics in comparison to a REP_benchmark. This can be seen as a gap in the literature. In the following paragraphs, different authors’ perspectives about portfolio and risk theories are briefly synthesised. The Subsection 3.3.1 defines the term "portfolio optimisation" with its relationship to the modern portfolio theory; the Subsection 3.3.2 presents the REP optimisation models; and the Subsection 3.3.3 discusses REP Systematic and Idiosyncratic Risks.

3.3.1 Portfolio Optimisation and Modern Portfolio Theory

Portfolio or asset allocation optimisation is defined by Sing and Ong (2000, p. 213) as '... the process of mixing asset weights of a portfolio within the constraints of an investor's capital resources to yield the most favourable risk/return trade-off.' For typical risk-averse investors, an optimal combination of investment assets that gives a lower risk and a higher return is always preferred (Markowitz 1952). Regarding the RE risk/return ratio and the RE diversification properties, Chapman (1979, p. 306) states that the '…Markowitz mean-variance analysis (MVA) approach to capital allocation seeks to maximize expected portfolio return on investment while minimizing the variability (risk) of this rate.' Chapman (1979) notes that, in addition to providing a model for incorporating risk, Markowitz provided a mathematical approach for the investor to systematically examine the risk/return trade off. This model, with its efficient frontier, is often adopted for GIP optimisation.

Although the above portfolio optimisation definition is valid for the REP, too, Coleman (2005, p. 38) notes that '…while this theory tends to work surprisingly well for highly liquid public markets, such as common stocks and corporate bonds, RE is another matter.' He remarks that the RE market does not conform well to many of the key assumptions underlying standard mean-variance optimisation because firstly, asset-returns are not normally distributed and secondly, the investor risk aversion is ignored. A confirmation of this fact is provided by Georgiev et al. (2003). They indicate that the RE return distributions are of importance for the portfolio manager as they provide key inputs into the asset allocation process.

According to Georgiev et al. (2003), much of the research has focused on testing for normality in RE returns and generally, in terms of skewness and kurtosis, normality is rejected both domestically and internationally and for both the direct and indirect market. Coleman's position can be seen in contrast with Chapman's opinion about the Markowitz model. In fact, Coleman states that the efficient frontier approach is not ideal for the REP optimisation. These authors' statements about the problems with the assumptions underlying the MVA confirm the need for the development of an additional REP evaluation system which is independent from RE returns and able to evaluate a REP from another point of view.

The MVA model, also referred to as modern portfolio theory (MPT), demonstrates how rational investors will use diversification to optimise their portfolios and how a risky asset should be priced. The MPT states that assets should not be selected solely on expected return and risk on each asset, but also on the correlation of returns for each and every pair of assets such as shares, bonds, cash, housing and other property. Thus, taking these co-movements into account it is possible to form a portfolio that has the same expected return and less risk than a portfolio built by ignoring the interactions between assets. The model assumes that investors are risk-averse and this means that given two assets that offer the same expected return, investors will prefer the less risky one. Thus, an investor will take on increased risk only if compensated by higher expected returns (Connor & Korajczyk 2009; Elton & Gruber1998; Sharpe 1964).

The limitations of the MPT, due to its assumptions, are described into the post-modern portfolio theory (PMPT) which, using recent advances in portfolio and financial theory, has lifted several of the MPT's limitations. PMPT shows how rational investors will use diversification to optimise their portfolios, and how a risky asset should be priced, thus the MPT is a special case of the PMPT (Zimmermann 2003; Elton & Gruber1998; Rom & Ferguson1994). Despite the contrasting researchers' statements, there is no clear evidence that a theoretical application of the MPT model cannot be used in RE market and that the RE market does not follow the traditional portfolio theory. In brief, the portfolio theory paradigm can be applied as basis for this study on the RE market.

Fama (1970) affirms that the extent to which a technical trading strategy, such as momentum or contrarian strategy, is capable of generating excess returns is an indicator of market imperfections. Starting from this position, Marcato and Key (2005) test the efficiency of the private RE market through a study that tests whether the momentum strategy for a REP can outperform a buy-and-hold strategy, i.e. a capital-weighted index such as the IPD index of the British market. The results of the study demonstrate that momentum strategies have the potential to generate significant excess returns in direct RE investment, although these potential gains are likely to be partially offset by the additional trading costs required to implement such a strategy into such an illiquid RE market.

Because high transaction costs combined with liquidity problems reduce the power of the application of mechanical strategies for optimisation of REP into the RE market, it is advantageous to have a REP that is close to the REP_benchmark. Having a REP that replicates the REP_benchmark could help to improve liquidity and reduce transaction costs. This is in part due to shorter advertising periods generating lower costs. Therefore an implementation of a mechanical or technical strategy into a managed RE portfolio could become more attractive. The research completed by Marcato & Key (2005) confirms that the RE market is still inefficient and indicates the importance for firms to have a REP_benchmark in order to better analyse their own REP. Developing this benchmark is the aim of this study.

The importance of including RE in a multi-asset portfolio is acknowledged by Lim *et al.* (2008, p. 93) stating that '…RE investment is now part of the wider capital market flows and a key component on any investment strategy.' They note that large real estate investors are looking to diversify their portfolios and the most commonly used rationale for including RE within a mixed-asset portfolio has been its low correlation with the other asset classes with the effect of reducing portfolio risks. RE cannot be considered in isolation, but must be placed in the context of other investment opportunities, notably equities and bonds. Location characteristics i.e. micro, macro and structural are becoming increasingly important with regard to the competition for development and investment opportunity.

Lim *et al.* (2008) note that, as shown from many studies, key criteria in the investment decision-making process and for determining allocations across the asset classes are expected return, risk and diversification benefits and that '…for RE to compete and survive as asset class, the need to develop the infrastructure to achieve this through indexes of asset class performance and benchmarks are important considerations' (Lim et al. 2008, p. 95). They conclude that the availability of reliable RE performance indicators is a key advantage. The research completed by Lim *et al.* (2008) accentuates the importance of RE as an asset class in a portfolio and the importance of the presence of benchmarks for comparative analysis, justifying the development of a REP-EDM.

Topintzi *et al.* (2008) define direct RE investments as those that include the investment and management of actual tangible RE. These investments are chosen as a good alternative investment due to their relatively high returns and low risk, but at the same time, are plagued by illiquidity and high costs, such as transaction, taxes and related costs. On the other hand, indirect RE investments, often defined as pooled or securitised investment vehicles, are particularly popular with investors due to their high levels of liquidity, transparency and diversification. They assert that the increase in investors' requirements for RE market measures has occurred for various reasons: a) The need for greater certainty about market performance, volatility and risk, b) the increasing need for RE benchmarks in order to perform comparative analyses against other REP or other well measured asset classes such as securitised RE, bonds and equities, c) the need to better understand the RE market, d) the desire to understand the appropriate allocations to RE within a multi-asset portfolio, e) the requirement to establish the true diversification benefits of RE strategies, and f) the growth in the property derivatives market has started to address major concerns of institutional investors relating to the illiquidity of much RE investing. The research undertaken by Topintzi *et al.* (2008) confirms the importance of the development of RE benchmarks and the need for more transparency in the RE market.

In his research Montezuma (2004) includes the following three analyses that are relevant for this study:

1) The institutional allocation of investment towards RE may be justified on two main financial grounds: a) Residential property is a more effective hedge against inflation than both shares and bonds; and b) unsecuritised RE investment not only generates risk-adjusted returns comparable to those of bonds and shares, but also exhibits low levels of correlation with classic assets groups, improving the diversification benefit in a mean variance Markowitz framework. He states that the empirical evidence about the ability to hedge inflation of unsecuritised RE has not been always consistent. Possible explanations for this inconsistency could be: a) The shortcomings of standardisation in the RE measurements, namely the lack of reliable time series data on RE prices and RE returns and b) the limitations of the modern portfolio theory framework that has been used on the empirical studies. In brief, according to the empirical literature and the theoretical arguments offered, the assumption that RE is a perfect short and long-term hedge against inflation cannot be validated. Nevertheless, there is strong empirical evidence that RE returns tend to be positively correlated with inflation and that RE is a more effective hedge against inflation than both shares and bonds.

2) The use of the MPT by pension funds could be problematic. The pension fund equation determines that the sum total of present assets, future contributions, and future asset returns must equal present and future liabilities and he affirms that the classic mean variance framework ignores the presence of liabilities in the decision process. This is not completely realistic since one of the major institutional investment policy objectives is to ensure sufficient assets to meet liabilities. In other words, the institutional investors must also tailor their asset holdings to hedge their liabilities. Accordingly, the maximisation of risk-adjusted future surplus value, that is assets minus liabilities, can imply that pension funds' allocations are different from those suggested by the MPT framework. Thus, the institutional allocation can be best seen in an asset-liability context, where the net wealth portfolio respectively the present value of future liability obligations minus present value of asset holdings is optimised, rather than in an asset-only context.

3) The relationship between the space market and the capital market influences the risk premium of direct RE investments. Montezuma (2004) affirms that this relationship depends on the level of integration among the two markets, which is determined by the extent to which assets in these markets are affected by common economic factors. When the two markets are significantly integrated, it is expected that a large asset substitution will occur and this will have an impact on the RE prices. Thus, whether the space market and capital markets are segmented or integrated and to what extent they are causally related, have implications for effective portfolio diversification strategies. MPT suggests that the greater the degree of integration of markets, the less important are the benefits from diversification, because the RE price changes would

be more closely and promptly related to the general economic fundamentals. However, because the RE market is not completely efficient and liquid, profitable arbitrage opportunities for multi-asset portfolio exist. Thus the investors can exploit differences in the risk-adjusted returns (price of risk) across different asset classes e.g. RE and shares until price discrepancies are corrected. Briefly, in practice, the RE assets offer attractive diversification opportunities for multi-asset portfolio.

In his research, Montezuma (2004) highlights the importance of the RE asset class in a portfolio, the problem due to RE market inefficiency and illiquidity and the problem in the application of the MPT by pension funds. This adds further to the case for the development of a REP-EDM that increases transparency in the RE market. In summary, whilst the MPT is theoretically usable for RE market, practical difficulties in the use of this model are reported by many researchers. These difficulties with the practical usability of the theoretical MPT model have pushed researchers towards the development of special RE optimisation models such as those described in the following section.

3.3.2 REP Optimisation Models

Most of the REP optimisation models, independent of the applied optimisation approach, try to optimise the risk/return ratio, basing the analysis on the internal parameters of the REP itself. Benjamin et al. (2001), for example, divide the REP optimisation into five categories: 1) risk and returns, 2) diversification and portfolio optimisation benefits, 3) returns on real estate versus other investments, 4) Real Estate Investment Trusts (REITs), and 5) inflation and real estate returns. Apgar (1995) proposes a scorecard model based on five factors that managers can use to evaluate their current RE situation. Starting with an estimation of the factors (amount, price, grade, area and risk) the managers receive a quick means of judging their RE's effectiveness. In this model, the factor risk includes an estimation of the systematic and risks such as credit risk, liquidity risk, operational risk and market risk.

Another REP optimisation model developed by Stoy and Schalcher (2005) studies occupancy costs, which encompass all recurring direct costs for buildings and the associated structures and land, whether they are incurred on a regular or irregular basis, from the time the building is useable until its demolition. The amount of occupancy costs studied, using multiple univariate regression analyses, depended on the following factors: Strategies e.g. maintenance strategies, building characteristics e.g. standard and condition of building services, location e.g. compensation level of the region and usage e.g. usable floor area / existing work space. Although the model developed by Stoy and Schalcher's (2005) considers various RE characteristics it does not include a REP_benchmark as done in this study.

The REP optimisation models deal with the financial economics that are concerned with how capital markets weight the timing, risk and other attributes of the possible future cash flows from different types of assets to determine what these assets are worth in the market today; that is, at what price they trade or could be traded. Investors may differ not only in their investment time horizons, but also in their preferences for risk-taking in their investments and portfolios. These different models focus on the relationship between benefits or returns, costs and risks and, although they include the RE characteristics in their analysis, they do not integrate the physical characteristics of RE market with a real empirical REP_benchmark.

3.3.3 REP Systematic and Idiosyncratic Risks

The risk of a REP comprises *systematic risk*, also known as non-diversifiable risk that refers to the risk common to all RE, i.e. market risk. *Non-systematic* risk, also referred to as diversifiable risk or idiosyncratic risk, is the risk associated with individual RE. According to the modern portfolio theory (MPT), a rational investor should avoid idiosyncratic risk because only non-diversifiable risks are rewarded from the market. In other words, the idiosyncratic risks are not priced and should be avoided (Zimmermann 2003).

Zikmund (2003, p. 564) defines the '…beta as the appropriate measure of the systematic risk' of a portfolio. He adds that the tendency of a specific portfolio to move with the market is reflected in its beta coefficient, which is a measure of the portfolio's volatility relative to an average portfolio that tends to move up and down in step with the market. This average portfolio will, by definition, have a beta of "1" which indicates that if the market moves up by ten percent, the average portfolio will also move up by ten percent. A portfolio with a beta of 0.5 is only half as volatile as the market, it will rise and fall only half as much and it would be half as risky as a specific portfolio with a beta of "1."

Crouhy *et al.* (2001) divide the major financial risks that affect most investors into the following five categories: a) Market risk that is subdivided into equity risk, interest rate risk, currency risk, and commodity risk, b) credit risk that is decomposed into transaction risk and portfolio concentration risk, c) liquidity risk that is subdivided into funding liquidity risk and *trading-related liquidity risk*, d) operational risk that is decomposed into human factor risk, systems risk, model risk and technology risk, and e) legal and regulatory risk. In the RE direct investment market, the *trading-related* '…*liquidity risk* is potentially a major constraint or concern in RE investment because property assets can require a long time to sell at full value, compared to publicly traded stocks' (Geltner & Miller 2001, p. 130). The low liquidity, combined with an inefficient RE property market, increases the investor's decision difficulties to make a decision on its RE investment. Increasing the market efficiency, respectively the information available about the RE market through a REP_benchmark, like a measure of the market similar to beta of "1" for a portfolio, helps the investor not only in making the decision about the RE investment but also in better managing the idiosyncratic risks i.e. the trading-related liquidity risk.

For example, if an institutional investor such as a pension fund manager knows where the reference of the RE market is through a REP_benchmark of all pension funds in its metropolitan statistical areas (MSAs), then he can determine with the REP-EDM the "distance" between the RE investment that he wants to make and the REP_benchmark. In other words, he can determine if the RE investment lies "near" or "far away" from the REP_benchmark and, with this information and according to his strategy, make the decision whether to undertake the RE investment.

In this example, the information about the "distance" between the RE investment and the REP_benchmark increases the transparency of the planned RE investment. The

REP_benchmark can be seen as the reference point that indicates how the portfolios of all other RE market players in the pension fund industry are composed and therefore it gives an indirect indication about the trading-related liquidity risk. If the RE investment lies “near” the REP_B, then it should be easier to sell a property because it corresponds to what the RE market holds at this moment. Liquidity risk should be lower. If the RE investment lies “far” away, the opposite might be true. Having this information, the investor has more information upon which to base a decision. Given the importance of RE physical characteristics for RE investments, the REP_benchmark of these characteristics is likely to provide valuable additional information to the decision-making process.

The REP_benchmark may be viewed as an indicator of the “position” of the RE market. It indicates the location of the highest trading-related liquidity at a specific moment and this corresponds to the interaction of the supply and demand on the RE market. The market position in terms of price for a specific quoted share can be followed real-time at the stock exchange. The same cannot be affirmed for the RE market because it does not have a real-time pricing system. Thus a REP_benchmark can help to find the position in terms of “distance” between a firm’s REP and the REP_benchmark, basing the analysis on the physical characteristics of the RE market. According to the modern portfolio theory (MPT) a replication of the RE market would reduce the idiosyncratic risks.

In summary, although the MPT is characterised by limited practical usability for RE investment, there is no clear evidence that the RE investment cannot, in principle, be guided by modern portfolio theory. The importance of the RE asset class in a portfolio, the problems due to RE market inefficiency and illiquidity and the problems with the application of the MPT by pension funds justify the development of a REP-EDM. This study develops a REP-EDM that offers a measure in terms of "distance" of trading-related liquidity risk using the physical characteristics of the RE market with a real empirical REP_benchmark. This increases the information available on the RE market.

3.4 Benchmark: Background

The following subsections are built with a top-down approach. The Subsection 3.4.1 introduces the origin and the importance of benchmarking presenting a discussion through the relationship between quality, continuous improvement and benchmarking. The Subsection 3.4.2 presents various definitions of benchmarking highlighting its common characteristics and the relationship between the words "benchmarking" and "benchmark." The Subsection 3.4.3 presents a discussion of benchmarking classifications. The Subsection 3.4.4 develops a picture of different process models. The Subsections 3.4.5, 3.4.6, 3.4.7 and 3.4.8 present discussions of possible measurements that can be used to build a REP_benchmark.

3.4.1 Origin and Importance of Benchmarking

Douglas (1998) argues that total quality management (TQM) theory is the most all-encompassing and integrating approach to information-age organisations. He affirms that TQM is based on the system theory including the management of strategic systems that is the integration of all value-adding functions and processes, in an effort to continuously improve products and services on behalf of customers. Although each company must find its own way to implement its TQM, continuous improvement is a common denominator and a major component of any quality system (Douglas 1998; Elmuti & Kathawala

1997). As an example, one of the fundamental concepts in the European Foundation for Quality Management model is continuous learning and improvement (EFQM 2008).

Many tools and methodologies for measuring and improving business performance have been developed. ‘One such method, widely regarded as one of the most effective methods, is benchmarking. Benchmarking is considered one of the most effective continuous improvement tools of transferring knowledge and innovation into organizations’ (Jain et al. 2008, p. 102). Benchmarking is a key tool in the best practices of management and occupies an important place in total quality management paradigm (Jain et al. 2008; Zairi & Whymark 2000). For a company, continuous improvement means knowing where and what can be improved and knowing if its performance is better than that of its competitors. Benchmarking can help in answering these questions.

Benchmarking was launched and developed in the 1980s by the Xerox Corporation as a performance improvement system (Zairi & Whymark 2000; Fong et al. 2001). According to Douglas (1998), benchmarking has become an important environmental feedback mechanism for most information age organisations. Because information can be considered to provide a competitive advantage, benchmarking is a critical tool for learning organisations enhancing improvement of the performance. Douglas (1988) asserts that benchmarking assists an organisation in determining strengths and weaknesses and that it is fundamental to any continuous improvement initiative. Thus, when benchmarking is used to support continuous improvement strategies, it has a positive impact on competitiveness (Riberio and Cobral, 2006). In other words, benchmarking is an instrument for competitive analysis and performance evaluation.

In today's environment, change is the only constant and '…some organisations, like the various creatures of the world, will learn and adapt while others will become extinct' (Douglas 1998, p. 10). For the European Foundation for Quality Management, too, benchmarking is a critical factor. In fact the EFQM says that '…excellent organisations continuously learn, both from their own activities and performance and from that of others. They rigorously benchmark, both internally and externally' (EFQM 2008, p. 6). Douglas (1998) adds that benchmarking provides: a) A holistic systems perspective including best practices as comparison, b) information regarding gaps in performance in relationship to other organisations, c) motivation for improvement when best practices of others are superior to one's own, d) early insights into breakthroughs discovered by others which could be of benefit to the organisation, and e) input to the collective "profound knowledge" of the organisation.

Alstete (2008) acknowledges the importance of the benchmarking. He adds that the benchmarking process does not stop at the end of a project, but continues throughout the life of the organization. 'Comparing performance might give a company some new ideas and improve business for a while but in the end it might just be putting a band aid on the real issues that could have been worked out if the company had gone through each step of the benchmarking process' (Alstete 2008, p. 185). Elmuti and Kathawala (1997, p. 242) states that '…any company should do benchmarking if it wants to attain world-class competitive capability, prosper in a global economy, and above all if it wants to survive.' In brief, when benchmarking is used to support continuous improvement strategies, it has a positive impact on competitiveness and it can be a source of competitive advantage

(Riberio & Cobral 2006; Sarkis 2008). These various authors' statements highlight the importance of benchmarking in the business world and this study focuses exactly on the building of a REP-EDM model including a REP_benchmark which can be incorporated into the process of continuous improvement.

3.4.2 Benchmarking and Benchmark Definition

Benchmarking means different things to different people. Although a number of researchers have explored this concept, no consensus on a single definition of benchmarking has been achieved. In the literature, numerous articles reveal a variety of interpretations regarding what the term "benchmarking" actually means among users and among organisations as a whole (Alstete 2008). In the following paragraphs, different authors' perspectives about benchmarking are summarised beginning with an overview of their definitions (Table 3.2) and concluding with a discussion that consolidates the main concepts.

Table 3.2 Benchmarking Definitions in the Literature

Literature	Definition
Camp (1989)	Benchmarking is the search for the best industry practices that will lead to superior performance
Camp (1992)	Benchmarking is the continuous process of measuring products, services and practices against the company's toughest competitors of those companies renowned as industry leaders (Alstete 2008)
Vaziri (1992)	Benchmarking is the process of continually comparing a company's performance on critical customer requirements against that of the best in the industry (direct competitors) or class (companies recognised for their superiority in performing certain functions), to determine what should be improved (Fong et al. 2001)
McNair and Leibfried (1992)	Benchmarking is an external focus on internal activities, functions, or operations in order to achieve continuous improvement
Spendolini (1992)	Benchmarking is a continuous, systematic process for evaluating the products, services and work processes for organizations that are recognised as representing the best practices for the purpose of organizational improvement
Watson (1993)	Benchmarking is a continuous search for, and application of significantly better practices that lead to superior competitive performance
Partovi (1994)	Benchmarking is the search for the best industry practices, which will lead to exceptional performance through the implementation of these best practices
Elmuti and Kathawala (1997)	Benchmarking is the process of identifying the highest standards of excellence for products, services, or processes, and then making the improvements necessary to reach those standards, commonly called "best practices". The justification lies partly in the question: "Why re-invent the wheel if I can learn from someone who has already done it?"
American Productivity and Quality Centre (1998)	Benchmarking is the process of improving performance by continuously identifying, understanding (studying and analyzing), and adapting outstanding practices and processes found inside and outside the organization and implementing the results
Falk (2000)	Benchmarking is a continuous improvement process during which processes and methods of operational functions as well as products and services of one's own company are measured against a benchmark, i.e. the maximum achievable performance
Bhutta and Huq (1999)	Benchmarking is first and foremost a tool for improvement, achieved through comparison with other organizations recognized as the best within the area
Kumar et al. (2006)	Benchmarking is the process of identifying, understanding, and adapting outstanding practices from organizations anywhere in the world to help an organization improve its performance. It is an activity that looks outward to find best practice and high performance and then measures actual business operations against those goals (Anand & Kodali 2008)
Anand and Kodali (2008)	Benchmarking is a continuous analysis of strategies, functions, processes, products or services, performances, etc. compared within or between best-in-class organisations by obtaining information through appropriate data collection method, with the intention of assessing an organisation's current standards and thereby carry out self-improvement by implementing changes to scale or exceed those standards

(Source: Adapted from Jain et al. 2008, p. 104)

Although different definitions of benchmarking exist within the literature, they all share the same themes. In fact, all these definitions highlight common characteristics such as the continuous improvement process, search for best practices and systematic process of comparison (Jain *et al.* 2008). The word "benchmarking" includes a continuous improvement process and should not be confused with the word "benchmark" that is defined by Alstete (2008, p. 179) as '... a point of reference from which measurements may be made ... something that serves as a standard by which others may be measured or judged.' Gleich *et al.* (2008) add that it is essential that a benchmark is built with a quantifiable measure and that it is included in a performance measurement system in order to be able to perform a comparative analysis under the same conditions and settings as a part of the benchmarking process. This difference is relevant for this study. In fact, its focus is in building a REP_benchmark that can be integrated into a benchmarking process of a firm.

3.4.3 Benchmarking Classification

In the literature there is no evidence of a definitive consensus about the classification of benchmarking. Various authors have categorised benchmarking according to different criteria such as aim, focus and the bases of comparison (Anand & Kodali 2008). Watson (1993) suggests that benchmarking has undergone five generations: Reverse engineering, competitive benchmarking, process benchmarking, strategic benchmarking and global benchmarking. During this evolution, benchmarking was classified and re-classified and the various forms of benchmarking are not necessary mutually exclusive but rather complementary.

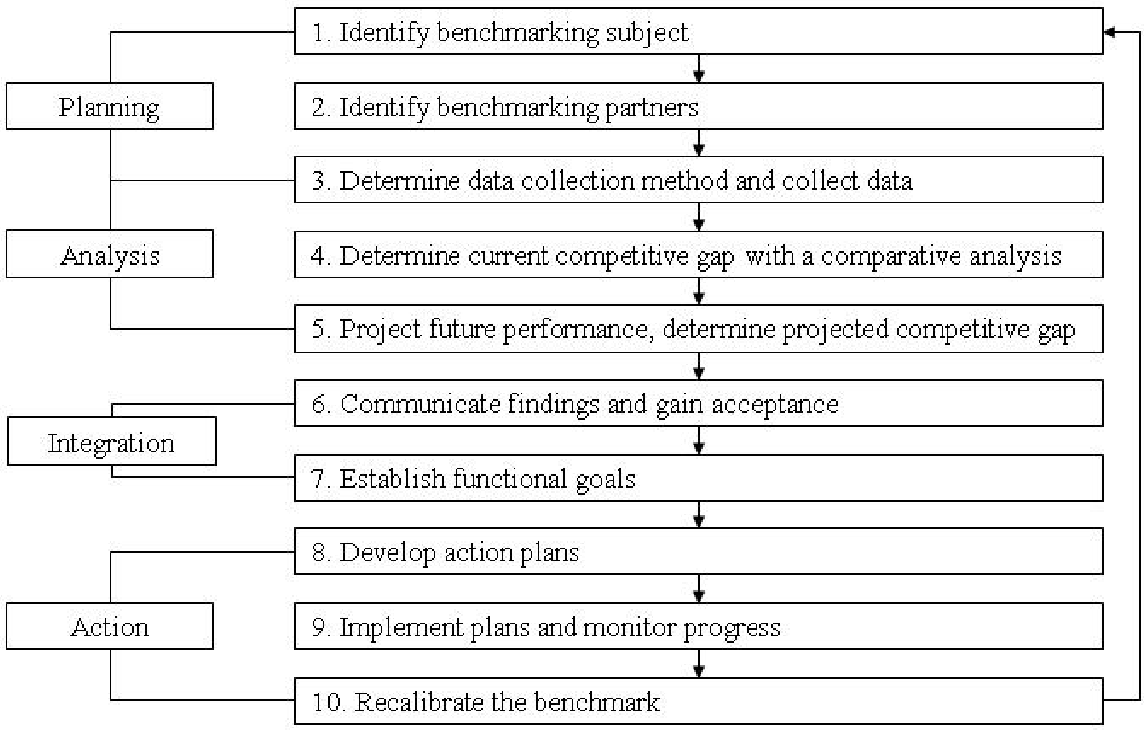
Anand and Kodali (2008), after an analysis of thirty-five published benchmarking models, suggest that benchmarking should be classified as: a) *Internal benchmarking,* done within the organisation allowing as example departments to learn from each other, and b) *external benchmarking* that is done with the external world. This can be, for example, searching for competitors who excel at a given process the organisation may want to improve, building on a related process in the same industry but with a firm with which the organisation does not directly compete, comparing with recognised leaders in the field which perform this process better than any other or, as in this study, using a specific external REP_benchmark in order to perform a comparative analysis. All other cases, like competitive, strategic, process, functional, etc. can be listed under these two categories. Such a classification scheme for benchmarking is simple and can reduce confusion among the practitioners.

3.4.4 Benchmarking Process Models

According to Anand and Kodali (2008), the purpose of the benchmarking process models is to describe the steps that should be carried out while performing benchmarking. Although the core of various benchmarking approaches is similar, most of the authors have divided their models differently. In fact, some companies have used up to thirty-three steps while others have used only three. In addition to the most widely used Xerox pioneering ten-steps benchmarking process (Figure 3.2), at least other sixty different existing models have been developed and proposed by various academics, researchers, consultants and experts. This increases the difficulty for the practitioners when it becomes necessary to choose a particular model for benchmarking.

Anand and Kodali (2008) propose the following taxonomy and divide the models into three main types: a) Academic/research-based models, which are developed mainly by academics and researchers mainly through their own research, b) consultant/expert-based models, which are developed from personal opinion and judgement through experience, and c) organisation-based models, which were developed by organisations based on their own needs. They state that Xerox's benchmarking model is the most commonly used model by the practitioners because it is considered to be an effective and generic way of conducting benchmarking. A detailed view of the Xerox benchmarking model which includes four phases and ten steps is depicted in the following Figure 3.2.

Figure 3.2 Xerox Benchmarking Model



(Source: Camp 1989)

Douglas (1998) tries to further reduce the model's complexity stating that, independently of the benchmarking models, the benchmarking consists of the three major-steps: a) measuring the performance levels of best-in-class relative to strategic process or performing a comparative analysis with the market, b) analysing how performance levels of these processes are achieved or analysing the gap to the market, and c) developing an improvement initiative using what has been learned.

The aim of benchmarking in RE investments is to enable an organisation or investor to learn from the best properties or portfolios by using the differences identified between properties or portfolios as the basis for developing plans for the improvement of one's own competitive edge (Falk 2000). Thus, this study deals with the Douglas' first two major-steps and Camp's first four steps (Xerox benchmarking model), in which a REP_benchmark for PFs is developed, a measure to calculate the PF's REP "distance" or to the REP_benchmark is defined and an analysis of the results is undertaken. PFs can then use the REP_benchmark to perform a comparative analysis in order to increase understanding and transparency of their REP, thus supporting strategic decisions.

3.4.5 Benchmark Measurements

According to Falk (2000) a "benchmark" in every-day language is a point of reference for a measurement. It can be seen as a standard against which something can be measured or assessed. A benchmark can be built with any basis of measurement, such as an interest rate, an index or peer grouping of stock or bond prices or other values, used as a reference point. Thus, it can be a simple measurement or a calculated result undertaken using different methods. For example it can be calculated as a weighted (*w*) average of a number (*i*) of the criteria (*c*) at a specific time (*t*) as in the Formula 3.2.

$$Benchmark_t = c_t^1 w_t^1 + c_t^2 w_t^2 + \ldots + c_t^i w_t^i \tag{3.2}$$

According to Georgiev *et al.* (2003, p. 29) the academic research on RE investment focuses on three principal benchmark areas: 1) The risk/return analysis, including diversification properties of RE investment and RE asset allocation issues, both internationally and within specific markets; 2) the economic determinants of returns in RE, that include the exploration of the relationship between RE returns and fundamental economic variables such as gross national product, real interest rates and growth in consumption and that includes the forecasting of RE markets cycles; and 3) the indexing and measurement issues in the RE market. This last benchmark area is explored in more depth, due to its relevance to this study. Before starting with the description of the existing indexes and measurements available in the RE market, some general and theoretical considerations about indexes, measurements and benchmark are presented below.

The questions of what an index is and which are the RE measurement issues are answered first. Vanini et al. (2008) define an index as a number that, over time, indicates the development of the measured variable. An index can be used as a benchmark to measure and compare the performance of a portfolio; to create derivative products; and as the basis for research that studies the market development, for example to find out which are the price drivers of a specific market, deducing investment recommendations. Critical factors when an index is used as measurement are: a) The representativeness that depends on the sample size and on the type of property and the completeness for the specific market under analysis, b) the availability, replicability, reliability, transparency, consistency and robustness in the calculation model, c) a sufficiently long time series for significant analysis, and d) the consistency between representativeness and market size weights definition of the considered variables. For example, if a specific share has ninety percent weight and the rest of the stock exchange only 10 percent in a whole market, the market representativeness of that index is questionable (Vanini et al. 2008; Topintzi et al. 2008).

In the literature various methods exist to determine the weighting of the variables in an index. Two of the most well known methods are the price-weighting and the value-weighting index methods. The price-weighting index corresponds to the mean value of the considered variables e.g. share prices and the calculation is done with the following Formula 3.3.

$$Pindex_t = \frac{S_t^1 + S_t^2 + ... + S_t^i}{n_t} \qquad (3.3)$$

Where

$Pindex$	: is the index value, the mean of the shares' price
S_t^i	: is the price of the share i at the time t
n_t	: is the number of quoted firms, shares at the time t

In this model, how many shares the individual firms have is not considered. The n_t must be adjusted by share-splitting so that the index value remains unchanged before and after the share-splitting. In practice, the share price is reduced by a splitting, thus the denominator must be reduced in order to compensate the numerator reduction in value. In the price-weighting models, a change of one percent in an expensive share induces a stronger change in the index value than a change of one percent in a cheaper share. In this case, a problem could arise that the index value becomes distorted if fast growing companies often split their shares. In fact, the company weight in the index would decrease. Internationally important indexes basing their calculation on the price-weighting model are: The Dow Jones Industrial Average and the Nikkei Dow Jones Stock Average (Vanini et al. 2008).

The value-weighting index corresponds to an adjusted price-weighting index. It considers how many shares the individual firms have and the proportional calculation is done with the following Formula 3.4.

$$Vindex_t = \frac{S_t^1 * n_t^1 + S_t^2 * n_t^2 + ... + S_t^i * n_t^i}{S_{t-1}^1 * n_{t-1}^1 + S_{t-1}^2 * n_{t-1}^2 + ... + S_{t-1}^i * n_{t-1}^i} * 100 \qquad (3.4)$$

Where

$Vindex$: is the index value, the adjusted mean in percent

$S_t^i * n_t^i$: is the capitalisation of the firm i at the time t

$S_{t-1}^i * n_{t-1}^i$: is the capitalisation of the firm i at the time t–1

n_t^i : is the number of shares at the time t for the firm i

In this model, the main problem is that firms with a high capitalisation value have a bigger influence on the index than firms with a lower capitalisation value. Internationally important indexes basing their calculation on the value-weighting model are the Standard & Poor’s 500 (S&P 500), the New York Stock Exchange Index (NYSE) and the

NASDAQ-series. In Switzerland, the Swiss Market Index (SMI) and the Swiss Performance Index (SPI) use value-weighting indexes (Vanini et al. 2008).

3.4.6 Index as Measurement

Although indexes themselves are not perfect, indexes are useful measurement tools as benchmarks of the market and practically all of the major types of investment products have their indexes:

1) The share-indexes can be divided into two categories. The first category is share-indexes that are not dividend-adjusted as they go down in value after a dividend has been paid such as the SMI that include the 20 most liquid and big firms of the Swiss stock exchange (SWX). The second category is share-indexes that are dividend-adjusted such as the SPI that include all the shares of the SWX.
2) The bond-indexes for example the Swiss Bond Index (SBI) have a special problem due of the limited duration of the bonds.
3) The interest-rate-indexes such as the London Interbank Offered Rate (LIBOR) is the most common of benchmark interest rate indexes used to make adjustments to variable rate mortgages.
4) The credit-indexes such as the Credit Default Swaps (CDS) index is a credit derivative used to hedge credit risk.
5) The RE-indexes that with the two RE investments possibilities that the RE market offers to an investor, can be divided into two categories: a) Direct-RE-indexes, and b) indirect-RE-indexes (Vanini et al. 2008). These two RE-indexes are discussed in the following in more detail, because they are relevant for the RE measurement issue.

When the benchmark is an index tracking a specific segment of the market, the changing value of the index not only measures the strength or weakness of its segment but is the standard against which the performance of individual investments within the segment is measured. An appropriate stock or bond index can be used to gauge the performance of an investment such as a mutual fund. For example, the S&P 500 index is one of the most commonly used benchmarks for comparing the performance of stock portfolio managers. There are other indexes that serve as benchmarks for both broader and narrower segments of equities markets, of international markets and of other types of investments such as bonds, mutual funds and commodities (Encarta 2007; Vanini et al. 2008). 'In contrast to the stock market, direct real estate does not really have a truly passive index' (Geltner & Miller 2001, p. 729).

Direct-RE-indexes cannot be considered passive indexes in which allocations across assets remain constant, such as the S&P 500. In fact '…the only type of periodic return indexes available to serve as benchmarks in the direct RE asset class are peer group indexes' (Geltner & Miller 2001, p. 731). While a peer group index contains, in principle, all of the performances of a certain type of REs that are currently active, a potential survivorship-bias problem is often raised concerning the appropriateness of peer universes as compared to passive indexes for performance benchmarks in the securities industry. Survivorship-bias occurs when REs have been taken off the market and disappear from the peer universe. In addition, Marcato and Key (2005) acknowledge that it is not really possible for any investor to buy the index, because the property composition of the index changes over time and the properties in the index are not for sale all times.

3.4.7 Appraisal-based Index, Transaction-based Index and Indirect RE Index

Various *direct-RE-indexes* have been developed and most of these are based on one of two principles. The first principle includes an index that is composed of *appraisal-based independent valuations* of a sample of REs coming from a specific RE market. These indexes provide a measure of investment performance such as total return, appreciation, income return, market rental value growth and gross or net equivalent yields for property comparable with the standard measures of investment return for other asset classes such as equities and bonds (Lim et al. 2008).

The methodology used to construct the index suffers from a series of weaknesses, such as subjective valuation inaccuracy, return or valuation smoothing, volatility dampening and artificially induced seasonality (Georgiev et al. 2003, Booth & Marcato 2004). Examples of such indexes are the National Council of Real Estate Investment Fiduciaries (NCREIF) index in the United States and the Investment Property Databank (IPD) index in Great Britain which essentially reflect the property-level performance of a major part of the private RE investment managers in the country and can be used as an indicator of the real estate's long-term average realized investment performance relative to other asset classes (Fisher & Geltner 2000; Lim et al. 2008).

The IPD method of calculating annual returns has been standardised for all reporting worldwide upon a single time-weighted approach, conforming to international standards, and in particular, the Global Investment Performance Standard (GIPS) for RE. The IPD index is calculated through two steps, the first using the Formula 3.5.

$$R_m = \frac{CV_t - CV_{t-1} - CExp_t + CRpt_t + EN_t}{CV_{t-1} + CExp_t} \quad (3.5)$$

Where

R_m	: monthly total return
CV	: capital value
$CExp$	: capital expenditure
$CRpt$	: capital receipts
EN	: net rental income
t	: month end
$t-1$	: month beginning

The second step, in order to calculate the time-weighted annual return, is done with the Formula 3.6.

$$R_a = (1 + Rm_1)(1 + Rm_2)...(1 + Rm_{12}) - 1 \quad (3.6)$$

Where

R_m	: monthly total return 1 to 12
R_a	: time weighted annual return

Although IPD indexes provide useful insights and set the standard for measuring global performance, an additional weakness of the valuation-based return of RE is the lack of availability and the robustness of individual country series (Topintzi et al. 2008).

The second principle is based on an index that measures the change in property prices and it is built on a *RE transaction-based* (sell-buy) pool. It is normally calculated only if the number of included transactions reaches a representative number of samples for the population considered. Vanini et al. (2008) suggest that such indexes have the advantage of higher flexibility during market changes, respectively shorter reaction times. Using real transactions, the opinions of the investors from the supply and from the demand side are immediately reflected in the index. Examples of such indexes are: a) For commercial properties, the Russel-NCREIF, and b) for residential property, the international Halifax House Price Index, the S&P Case-Shiller Index and the ZWEX a residential property price index for the Canton Zurich (Figure 2.4).

Theoretical and empirical studies agree that a RE price-indexes can be calculated with a hedonic regression model that includes micro, macro and structural characteristics as predictor variables for the estimation of the price. Some of the most important variables are the distance to town centre, surface, condition of RE, lake view and taxation level (Studenmund 2006; Prioni & Bignasca 2005; Häussermann et al. 2004; ZKB 2004; Geltner & Miller 2001; Linneman 1980; Rosen 1974; Grether & Mieszkowski 1974). Hedonic regression decomposes the item being researched into its constituent characteristics and obtains estimates of the value of each characteristic. In essence, it assumes that there is a separate market for each characteristic; thus it is one of the best ways to measure the inflation of heterogeneous good (ZKB 2004; Freeman 1993). The hedonic multiple linear regression model is represented by the Formula 3.7.

$$E(RE) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i \qquad (3.7)$$

Where

$E(RE)$: represents the mean or expected RE price

β_{0-i} : are the regression's coefficients

x_{1-i} : represents the RE's characteristics (micro, macro)

After having analysed the measurement issues of the direct-RE-indexes, the other category, the *indirect-RE-indexes,* may be briefly considered. The problems with real or direct-RE-indexes might encourage the use of data from financial or securitised RE, for which market data is readily available. Examples of such indexes are: a) For the world, the National Association of Real Estate Investment Trusts (NAREIT) index and the Global RE Return Index, b) for Europe, the European Public Real Estate Association (EPRA) index of publicly traded REITs and RE funds, and c) for Switzerland, the SWX Real Estate Index, which contains all SWX-listed real estate investment companies, the SWX Immobilienfonds Index, which contains all SWX-listed funds and the Rüd-Blass Immobilienfonds Index, which contains a selection of the ten biggest SWX-listed funds.

Unfortunately, research has shown that indirect-RE-indexes are an inadequate representation of the underlying physical RE market. In fact returns on REITs are nearly uncorrelated with returns in the direct RE market, but rather are more closely related to equity markets (Georgiev et al. 2003). An additional problem, according to Booth and Marcato (2004), is that indirect-RE-indexes do not properly measure the value investors put on the underlying assets of RE companies because RE companies are geared towards profit. These problems legitimate the need for indexes based on RE physical characteristics.

In Switzerland, two of the major players offering RE-indexes, are the companies "Informations- und Ausbildungszentrum für Immobilien AG (IAZI)" and "Wüst & Partner AG (W&P)" (IAZI 2007; W&P 2007).

- The SWX IAZI Real Estate Index Family is divided into two transaction-based indexes: 1) The *private real estate indexes,* which include the SWX IAZI Private Real Estate Price Index (SFD+CO), SWX IAZI Private House Price Index (SFD) and SWX IAZI Condominium Price Index (CO), and 2) the *investment real estate indexes,* which include the SWX IAZI Investment Real Estate Performance Index and the index SWX IAZI Investment Real Estate Price Index. In addition, IAZI provides appraisal-based indexes such as the IAZI Swiss Property Benchmark® that include end 2006 over 7,000 Swiss properties with a market value of circa 70 billion CHF. Supplementary benchmarks calculated by IAZI include the evolution of return, prices, vacancies, return per m^2 and cost per m^2 per canton, regions and cities (IAZI 2007).
- In parallel to IAZI, W&P provides also with two categories of indexes: 1) The *W&P RE Benchmark*, which includes direct-RE-indexes about the evolution of prices with drill down possibility to regional areas, and 2) the *W&P WUPIX Benchmark*, which includes indirect-RE-indexes that are share dividend-adjusted for RE companies and funds (W&P 2007).

Because of their independence from the banking world IAZI and W&P are considered to be third parties, which drive the transparency of the Swiss real estate market. Although they are the publishers of most of the periodicals about RE market present on the official Swiss Stock Exchange (SWX) web page, they do not offer a REP_benchmark based on the physical characteristics as done in this study.

3.4.8 Complementary RE Benchmark for the Unsecuritised RE Market

There are problems with the *appraisal-based indexes* such as subjective valuation inaccuracy, return or valuation smoothing, volatility dampening and artificially induced seasonality. There are problems with the *transaction-based indexes* in reaching a representative number of samples for the population to be considered. Further there are problems with the *indirect-RE-indexes*, which are classified as inadequate representations of the underlying physical RE market. Due to these issues, researchers continue to search for additional measurement systems that can be used as benchmark for the unsecuritised RE market. Thus, in this section, other or complementary methods of RE benchmark are presented.

3.4.8.1 Jones Lang LaSalle RE Transparency Index

Not only performance, returns or prices can be benchmarked. In fact, the Jones Lang LaSalle Real Estate Transparency Index is one of the benchmarks used by the international investment community to compare the *risk* profiles of regional and global property markets (Muller 2006). This index is built using a survey method and includes the following five attributes for the measurement of the RE transparency: a) availability of investment performance indexes, b) availability of market fundamentals data, c) listed vehicle financial disclosure and governance, d) regulatory and legal factors and e) professional and ethical standards (Mortgage Banking 2006). This index increases the RE market transparency with respect to the risks of a RE investment in a specific geographical area.

3.4.8.2 Single Measures as RE Benchmark

The benchmark proposed by Massheder and Finch (1998) divides the RE benchmarks into the following categories: a) Monetary benchmarks e.g. CHF per m^2 usable floor area and year, b) physical benchmarks e.g. kWh per m^2 usable floor area and year, c) productivity benchmarks e.g. percent, d) building efficiency benchmarks e.g. m^2 usable floor area per m^2 gross external floor area and e) capacity benchmarks e.g. m^2 usable floor area per existing workplace. An example of how these types of benchmark can be used is given by Stoy and Schalcher (2005) who collected and studied data using a survey done by four Swiss companies and which included 116 properties, the occupancy costs using building efficiency and capacity benchmarks as core indicators for the analysis. They compared the single measures or benchmarks to their analysed REP. These types of benchmark categories are used as a single measurement that is compared to the specific property under analysis but Stoy and Schalcher (2005) do not build a REP_benchmark that consolidates more physical characteristics of a specific RE market into a defined MSA. This is a gap that this study attempts to fill in developing a new type of benchmark based on physical characteristics.

3.4.8.3 Agents, Brokers and Organisations Data as RE Benchmark

A category of the suppliers of benchmarks such as the advertising duration time for selling a properties, the supply as percentage of housing stock, list of properties to sell or to buy, etc. are various firms such as brokers and agents that publish their data into the internet. The internet enhances transparency and market efficiency and thus contributes to minimising search costs when looking for suitable properties (Schulte et al. 2005) but again, no REP_benchmark based on the physical characteristics of RE is available as done in this study.

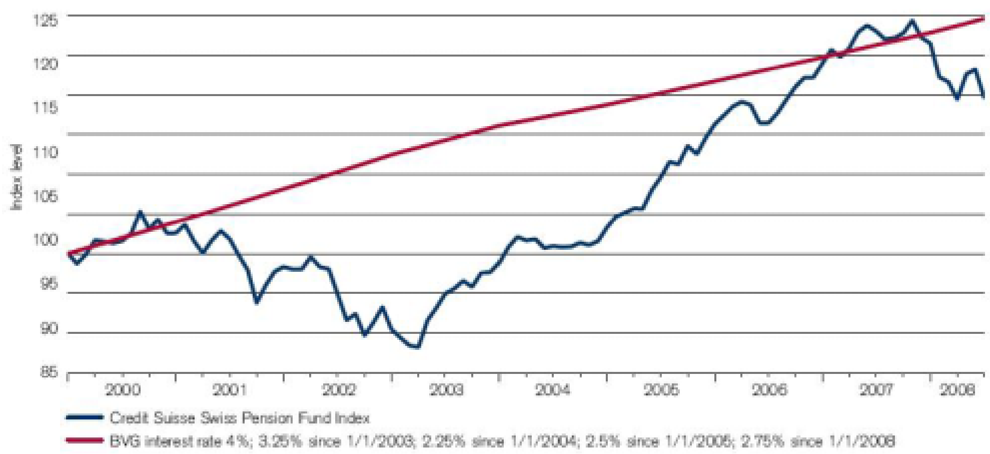
The need for a higher RE market transparency has pushed various organisations and companies to publish periodical market reports. Typically, a market report contains a review of the current state of affairs and an assessment of the general economic situation, as well as forecasts for future developments in residential, office and retail property. Example of measurements contained in such market reports and that can be used as benchmarks are: indexes, demand-side indicators, rents, net increase in homes, properties values and prices by cities and regions, vacancy rates, absorption rates, etc. (W&P 2007). Example of suppliers of periodical market reports and studies about the Swiss RE market are: Public institutions such as the statistical office of the Canton Zurich for the Canton Zurich RE market (STA 2008a) and the Swiss federal government for statistics (SFG 2008a) and private institutions or companies such as Zurich cantonal bank (ZKB 2008), Credit Suisse bank (Braun et al. 2008), Wüst & Partner AG (W&P 2007) and Informations- und Ausbildungszentrum für Immobilien AG (IAZI 2007).

These suppliers of benchmarks obtain their data through three major sources. The first source is the replies to questionnaires that are sent to different property owners. The second source is based on regular and uniform data collection from leading property owners and agents that have agreed to deliver the necessary data into a database at regular intervals. The third source is based on large properties-market players as well as public institutions that use the data from their own sources or of their own properties. Although these reports and studies help by increasing RE market transparency - they contain general views and trends for a region, for a canton and for the whole of Switzerland - they are not dedicated benchmarks for a specific segment of companies like the PFs and they do not build a REP_benchmark such as proposed in this study. Diverse suppliers of RE market reports, having realised that more precise and dedicated benchmarks are needed, have created their own specific benchmarks as described in the following.

3.4.8.4 CS Swiss Pension Fund Index

Credit Suisse (CS) provides a specific real indirect-RE-index, the "CS Swiss Pension Fund Index" which is published quarterly and it is based on the CS own managed capital, coming from various autonomous PFs, with a capitalisation of about 100 billion CHF (Figure 3.3).

Figure 3.3 CS Swiss Pension Fund Index



(Source: CS 2008)

When interpreting this index, it must be kept in mind that the CS Swiss Pension Fund Index is not an artificially constructed performance index but a real index that is based on actual pension fund data. The result is that the index is “alive”, which significantly increases its informative value, especially regarding the current investment behaviour of Swiss pension funds. The fact that it is regularly revised, however, limits the comparability of data over time (CS 2008).

3.4.8.5 KGAST Immo-Index

KGAST (Conference of Investment Foundation CEOs) is an organisation that provides the “KGAST Immo-Index” (Figure 3.4). This is a real indirect-RE-index and includes 25 RE investment foundations in Switzerland that are not traded on the stock exchange. In fact, their price tends to follow the performance of the real estate itself and it is less correlated by the development of the Swiss equities. KGAST contains 25 foundations as members and has a total capitalisation of about 72 billion CHF. A particularity of investment foundations is that they are only available to PFs that receive preferential tax treatment under BVG pension legislation (Watson-Wyatt 2008).

Figure 3.4 Performance of Swiss Real Estate Investment Indexes

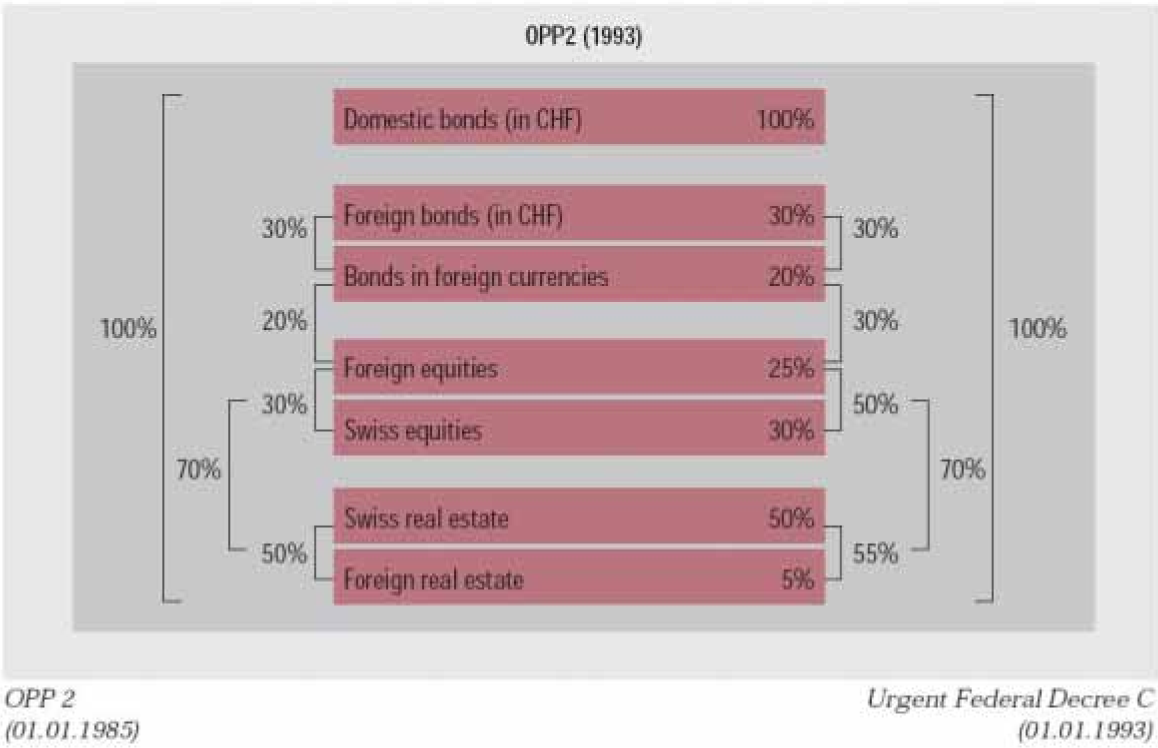
(Source: Braun et al. 2008)

Like KGAST, other organisations exist, such as ASIP-Swiss Pension Fund Association, Lusenti Partners AG, Complementa AG and Swisscanto AG that are specialised in PF performance comparison reports. These reports include different measurements such as indexes and risk (age-structure) adjusted performance that are built starting from various surveys among different PFs that agreed to participate to the inquiry.

3.4.8.6 Pictet-BVG-Index

In contrast to the CS and KGAST real indirect-RE-index, the Pictet-BVG-Index is an artificially constructed performance index. On January 1 1985, the law concerning the contingency fund (BVV2) came into force. This law standardised existing cantonal PF regulations regarding asset categories and their maximum authorised weightings. Simply measuring the profitability of funds invested does not suffice when it comes to judging the quality of management. Performance in itself is a function of the evolution of markets and currencies, as well as of the constraints laid down by the BVV2. For this reason, a performance measure is needed. Thus, Pictet & Cie bank calculates and publishes, since the first January 1985, on behalf of the Swiss Bankers' Association an index which measures a theoretical average performance of portfolios subject to the BVV2. The constraints imposed on pension funds by the BVV2, art. 53, deal mainly with 7 asset categories (Figure 3.5). They define maximum limits for each individual category and, on a broader basis, for combinations of several different investments (Pictet & Cie 1999).

Figure 3.5 The Investment Limits of the Swiss Pension Fund Index



(Source: Pictet & Cie 1999)

Figure 3.5 illustrates the maximum limits of investment given by the previous version of the BVV2 1985 (left) and by the present limits of the BVV2 1993 (right). The Pictet-BVG-Index is an average theoretical portfolio which respects the legal constraints. The index performance therefore measures a portfolio invested in a neutral manner within the legal prescriptions. In other words, the performance is equal to the average performance that all weighted portfolios would have realised including all authorised combinations. The Pictet-BVG-Index, which is recognised as an industry standard and is used as a benchmark to analyse PFs' return on securities investments, takes no account of RE and only partially incorporates cash assets (Pictet & Cie 1999). A first problem using this index as benchmark could be caused by the weight of the assets, which is fixed and does not necessarily correspond to the reality PFs asset allocation. The second issue is the absence in the index of the RE assets, which in the reality could be present in the PFs investments.

3.4.8.7 ZWEX - Residential Property Price Index

In the Canton Zurich the RE market is highly segmented and is a local market. This fact is also demonstrated through the high differences in prices of the empty ground available for building among its regions (Figure 2.3). The Zurich cantonal bank (ZKB) builds the ZWEX a residential property price index for the Canton Zurich (Figure 2.4). In order to depict a more precise measurement of the residential properties prices in accordance with the highly segmented RE market, the ZKB divides the ZWEX into two indexes: a) The ZWEX-See that include all municipalities of the canton that have contact to the lake, and b) the ZWEX-Regio that include the rest of the municipalities. In addition to these indexes, the ZKB provides specific analyses about each single region of the Canton Zurich. In fact, it supplies transaction-based price indexes for each single region and basing the analysis on 15 variables it calculates a rating for each region (ZKB 2008).

Collins et al. (2006) remark that the benchmark aspect of the benchmarking process is an area in need of further refinement. In fact, all the above described RE benchmarks increase the RE market transparency but no single one is based on a REP_benchmark that uses the physical distinguishing marks of the RE market. In order to address this limitation, this study develops a REP empirical decision model (REP-EDM) including a real REP_benchmark under consideration of the real estate physical characteristics. This model can be used as a decision support system by RE managers in the decision process that provides additional and detailed information regarding the RE market and that presents another point of view with an evaluation of the own REP situation.

3.5 Conclusion

This study is founded on existing work in the fields of real estate, finance and benchmarking theories and attempts to depict the situation in the current environmental context of direct RE investment. Benchmarking can be seen as a determinant of competitive advantage that increases the transparency of the RE property market, thus supporting investors in their RE investment decisions and promoting continuous improvement as part of the TQM paradigm. The development of the benchmark is a substantial part of the benchmarking process and it represents the both focus and the framework of this study.

The purpose of this study is to develop an objective REP-EDM based on a REP_benchmark based on physical characteristics of the RE market. The REP_benchmark can be used to perform a comparative analysis into a benchmarking process. According to the literature reviewed, most REP optimisation models are focused on the risk/return ratio, benefits, prices, and occupancy costs and there is no evidence of a REP-EDM that integrates the physical characteristics of RE market with a real empirical REP_benchmark. In brief, this research is justified in addressing a relevant topic of interest within the community that has not yet been empirically investigated and increasing the understanding, the body of knowledge and the transparency of the RE market.

4 RESEARCH METHODOLOGY

4.1 Introduction

Chapter One established the foundations for this research. Chapter Two presented Switzerland's RE market as a segmented market and outlined the participation of PFs in the market. Chapter Three presented the literature review aimed at establishing the importance of this study. This chapter identifies the appropriate research paradigm for this study. The research design and methodology used to collect and to analyse the field data addressing the described research issues is discussed and justified. The aim of this chapter is to provide assurance that appropriate procedures were followed. The chapter is organised around following topics: 1) Research paradigm, 2) research design, 3) research methodology including the data collection and analysis, 4) computation of the REP_benchmark, 5) development of the REP-EDM, 6) validation and reliability of the model, 7) limitations of the research, 8) ethical considerations and 9) conclusion.

4.2 Research Paradigm

Guba and Lincoln (1994, p. 107) suggest that '…a paradigm is a world view that seems so natural that it is accepted almost on faith' and it may be viewed as a set of basic beliefs about questions of reality, truth and objectivity that defines the nature of the world. The way data is perceived relates to the paradigm adopted. Leedy and Ormrod (2005, p. 88) describe '…data as manifestations of reality, with the truth underlying the data being unobservable.' They affirm that there are two barriers between the researcher and the absolute truth: a) The impenetrable barrier between truth and data; and b) the imperfect way a researcher perceives the data. In brief, the awareness regarding which is the most appropriate paradigm to guide the research is fundamental for the researcher who has to deal with different assumptions about how the world is perceived and how he can best come to understand it.

The relationship between the absolute truth and the researcher can be viewed under different paradigms that start from a most objective view and end with a more subjective view of the absolute truth. According to Leedy and Ormrod (2005) the positivist approach is used by quantitative research such as this study to answer questions about relationships between measured variables, using standardised procedures to collect numerical data and using statistical procedures to analyse and draw conclusions from the data. 'Almost all quantitative researchers rely on a positivist approach to social science' (Neuman 2003, p. 139). Table 4.1 outlines how the positivism is the appropriate paradigm for this study.

Table 4.1 Key Features of Positivism to this Study

Ontology	Question	What is the nature of reality?
	Response	This study of REP assumes that reality exists and that a RE is recognisable and understandable. This research investigates real situations of pension funds in Canton Zurich through an empirical design, in which evidence is rooted in objective reality.
Epistemology	Question	What is the relationship between the inquirer and that being studied?
	Response	A technocratic approach is applied in which the researcher is the expert that analyses existing data without being influenced by it and without being capable to influence it. In fact, the investigator can be separated from the object investigated. The objective is to analyse relationships between variables measured in the RE reality in order to solve the research problem eliminating the human factor and emphasising the objectivity of the developed REP-EDM.
Methodology	Question	How should the inquirer obtain knowledge?
	Response	For this quantitative research, the technique used to examine the reality is the correlational research method. Starting from existing data, that is collected from various sources, a REP-EDM is developed and validated through statistical analysis.

(Source: Developed for this research and adapted from Guba & Lincoln 1994)

4.3 Research Design

In order to establish the research design for this study, Neuman's (2003) four dimensions in research (how, purpose, time and data collection) are deployed to guide the research undertaken as part of this study as follows:

1) The first dimension defines how research is used: Applied vs. basic. This study uses applied research.
2) The second dimension defines the purpose of the study: Exploratory, descriptive or explanatory. This study is a descriptive analysis that provides a highly detailed,

accurate picture of the "distance" between the PFs real estate portfolio benchmark (REP_B) and a single PF's REP (REP_S). This descriptive research examines a situation as it is and it does not involve changing, manipulating or modifying the situation under investigation.

3) The third dimension defines the way time is incorporated into the study design: Cross-sectional, longitudinal (time series, panel, cohort) or case study. This study is a cross-sectional analysis.
4) The fourth and final dimension defines the technique for collecting data: Qualitative vs. quantitative. This study collects quantitative data in the form of numbers and this influences the research method that can be applied to solve the specific research problem, in fact '…to some extent, the data dictates the research method' (Leedy & Ormrod 2005, p. 94).

In summary, this study is an applied research project with a focus on the real estate portfolio (REP) management. The research project uses a cross-sectional design to obtain descriptive respectively quantitative data with the aim to develop a REP empirical decision model (REP-EDM) including the computation of the REP_benchmark for the pension fund (PF) RE market in the Canton Zurich. The REP-EDM allows the evaluation of a specific PF's REP (REP_S) under consideration of its RE physical characteristics and the pension fund REP_benchmark (REP_B). The developed REP-EDM can be used by PF managers to analyse and to optimise their respective REP_S compared to the REP_B.

4.4 Research Methodology

In this study, a PF REP_benchmark (REP_B) of the RE market physical characteristics is constructed and a measure of the "distance" between the REP_B and a specific single PF's REP (REP_S) is developed in order to identify the level of the divergence between the PF's REP and the benchmark. The principal focus is not to find out the causes and reasons of the current RE market situation, but rather to develop a REP_benchmark and determine a measure of the distance between the REP_S and REP_B. The procedure followed in this study, which is described throughout this chapter, consists of the following five steps:

1) Sampling design definition, data collection and consolidation, data analysis
2) Description of the variables used for PF REP characterisation
3) Computation of the PF REP_benchmark (REP_B) of the PFs in the Canton Zurich
4) Definition of the "distance" measure between a PF's REP and REP_B
5) Development and validation of the REP-EDM

4.5 Sampling Design, Data Collection

4.5.1 Population, Sample Selection and Sample Size

To reduce the problems associated with data validity, reliability and representativeness that may arise during the process of sampling the population, it was decided to focus the data collection for this study on the attainment of a complete population of data. This means that the entire PF population of the Canton Zurich as it existed on August 13, 2007, is included into the cross-sectional analysis. This date, August 13, 2007, corresponds to the delivery date of the data from the GVZ to the STA. In brief, the sample design for this study does not involve the selection of a technique to choose elements from the population of interest. No sampling of the population was undertaken. Rather a complete population of data was utilised.

4.5.2 Data Collection and Consolidation

In order to obtain a dataset of the PFs in Canton Zurich, the data collection procedure started with the request for data from the various sources as well as with the request for permission to use said data for this study. All sources generously provided approval to conduct this research and to access their database of PFs under the conditions defined in Table 4.2.

Table 4.2 Terms and Conditions to Obtain and Use Data

1) Data collection
The relevant data from GVZ (building insurance in Canton Zurich) were already obtained in the context of the register harmonisation by the statistical office and were stored on the secured database of the statistical office (STA). Consequently, no new GVZ-data needs to be delivered by the GVZ.
2) Data management
For the entire duration of the requested utilisation, the data remain in the statistical office within the protected area. Applicable to the intended study, data protection and data security regulations have to be assured, as with the GVZ-Data in the context of the register harmonisation.
3) Data treatment and editing
A match between the GVZ-data and the data furnished by the Swiss Federal Government for Statistics (SFG) is done. Consequently, a data subset of GVZ-Data is created (REs-ZH-Dataset) that contains all the PFs of Canton Zurich. This new dataset is stored and secured in the same way as the GVZ-Data in the context of the register harmonisation.
4) Anonymity
The anonymity of the newly created dataset (REs-ZH-Dataset) is assured, in that individually sensitive data such as the names and addresses of the pension funds are deleted once data has been aggregated and analysed. Consequently the database is anonymous and no link to individual PF or person can be made.
5) Data analysis
Starting from the physical RE characteristics such as example number rooms, surface, etc. a benchmark among all pension funds is calculated. As result, a fictitious and consequently completely anonymous “average pension fund” or “REP_benchmark” is determined. After that a statistical regression model is calculated, with which the “distance” can be calculated for a real PF’s REP to the REP_benchmark.
6) Data destruction after use
After the study is terminated, all used data is completely and irrevocably deleted. As a result, the fictitious “average pension fund” or “REP_benchmark” and the mathematical model remain.

(Source: Daniel 2008)

The raw data on REP from pension funds in Canton Zurich was collected systematically from the following sources:

a) GVZ - Building Insurance Zurich <www.gvz.ch>:

Source for building owners’ names and addresses for the Canton Zurich. The cross-sectional design is dictated by the GVZ and set on the August 13, 2007.

b) SFG - Swiss Federal Government for Statistics <www.bfs.admin.ch>:

   Source for PF names in the Canton Zurich.

c) ZKB - Zurich Cantonal Bank, Zurich, <www.zkb.ch>:

   Source of the GIS (geographic information system) data.

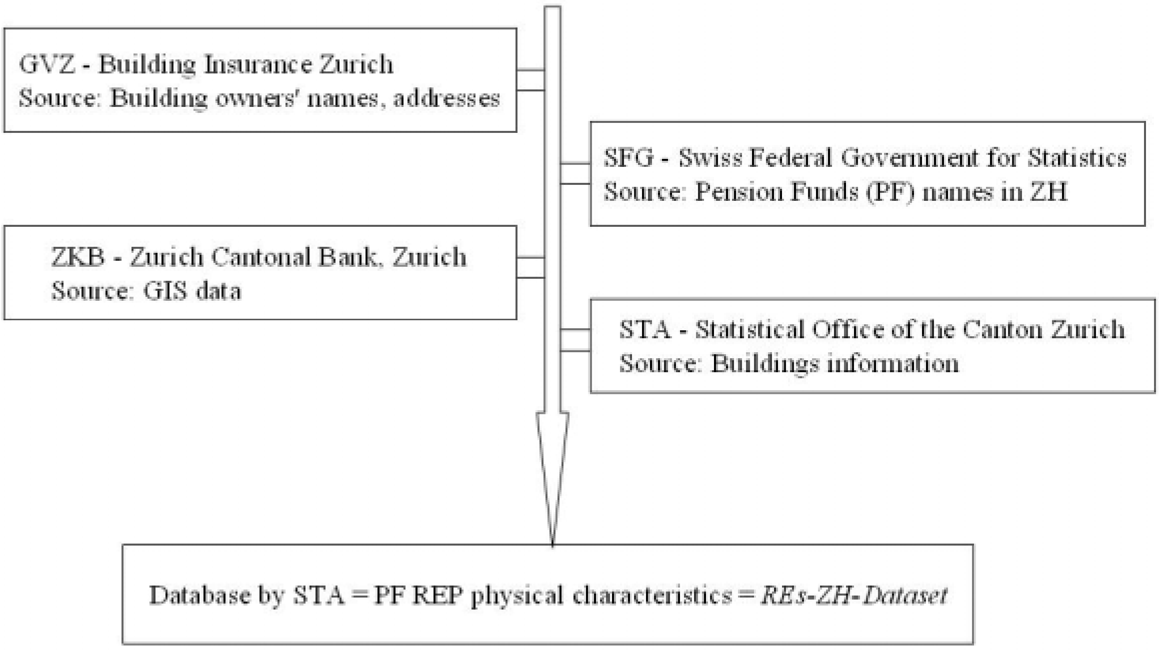
d) STA - Statistical Office of the Canton Zurich <www.statistik.zh.ch>:

   Source of the buildings' physical characteristics, geographical coordinates and major database for the data collection and consolidation.

e) For possible missing data, a collection of the data on the field is done manually.

The collected data, including structural, micro and macro-characteristics of the RE, are stored into the database provided by STA (Figure 4.1).

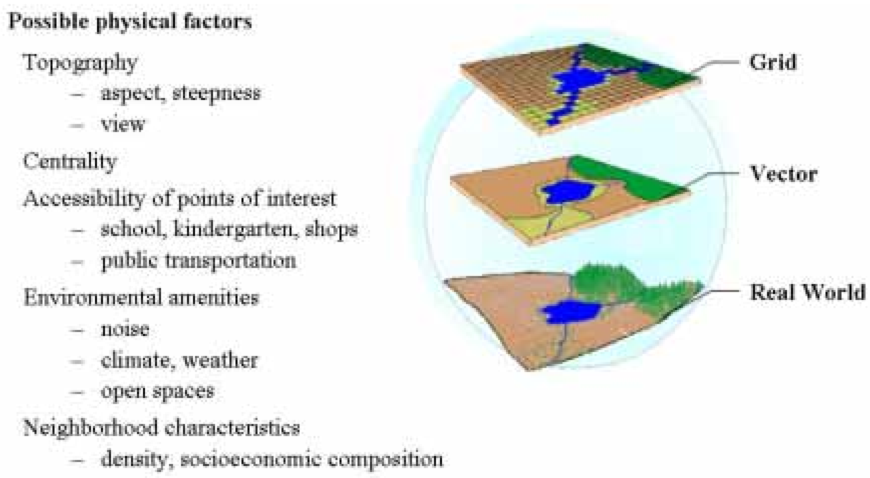
Figure 4.1 Data Collection



(Source: Developed for this research)

The data consolidation procedure starts by matching the PF names (SFG) with the building owners' names (GVZ) so that it can be determined which are the buildings that are owned by RE companies, in this case, the PFs. Diverse micro and macro-characteristics of the relevant buildings are determined using the geographic information system (GIS) technology of the ZKB. The 'GIS technology makes possible more geographically precise use of micro and macro information about the property' (Geltner & Miller 2001, p.113). The real world is converted into vectors and grids, which are used to calculate for example areas, distances and to map noise levels, etc. (Figure 4.2).

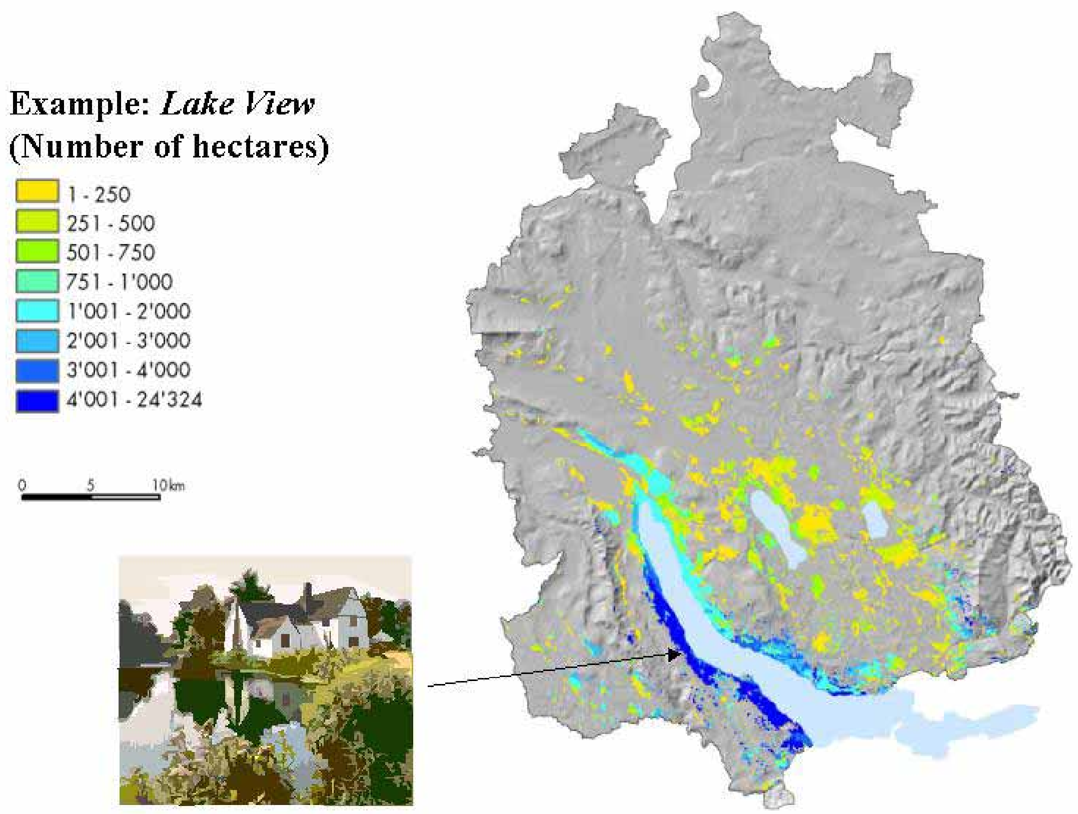
Figure 4.2 GIS Data Model



(Source: Developed for this research)

An example of the application of GIS is illustrated in Figure 4.3. In a first step, the topography of the Canton Zurich is decomposed in different layers calculating its attributes and various spatial information for each point on the grid. In a second step, the Lake-View area is computed from each point on the grid and the result is depicted with different colours. From properties in the yellow regions, 1 to 250 hectares of lake can be seen. From blue regions, up to 24,000 hectares can be seen.

Figure 4.3 GIS Example - Lake View



(Source: ZKB 2004)

This data represents an example of a physical characteristic that defines a RE and upon which the REP_benchmark is built. After the GIS micro and macro-variables have been calculated, the buildings' structural data, sourced from the STA, can be merged. The result is saved into a database called *REs-ZH-Dataset* which contains the RE physical characteristics such as usage, age, number of rooms, location and lake view for each RE owned by PFs that are located in the Canton Zurich. The structure of the REs-ZH-Dataset*,* filled with fictive data, is depicted in Table 4.3. An example of a possible content analysis is that the PF Nr. 1 owns eight REs, of which one is a single family dwelling (SFD) and seven are condominiums (CO).

Table 4.3 *REs-ZH-Dataset* with the PF REP Physical Characteristics

PF	RE	Structural Variables				Macro Variables			Micro Variables		
Nr	*Nr*	*Usage*	*Age*	*Nr of Rooms*	*etc.*	*Geo Location X*	*Geo Location Y*	*etc.*	*Lake View*	*Distance ToBus*	*etc.*
1	1	SFD	2	6	...	676924	237247		1500	200	...
1	2	CO	3	5	...	678524	238546		500	100	...
1	3	CO	6	4	...	676123	218562		0	50	...
1	4	CO	9	4	...	676841	236231		0	140	...
1	5	CO	15	6	...	685210	257357		0	234	...
1	6	CO	20	4	...	672354	237741		3000	122	...
1	7	CO	5	3	...	645632	227852		2500	342	...
1	8	CO	7	4	...	674582	237963		1200	23	...
2	1	SFD	8	6	...	676458	237258		0	200	...
2	2	CO	9	5	...	676742	231478		0	100	...
2	3	SFD	2	4	...	652136	242563		0	50	...
2	4	SFD	5	4	...	642156	285426		0	140	...
2	5	SFD	6	6	...	673251	225831		0	234	...
2	6	CO	14	4	...	673573	223556		0	122	...
2	7	CO	25	3	...	665423	233698		0	342	...
2	8	CO	68	4	...	652135	231478		2500	23	...
3	1	SFD	2	6	...	695213	238963		3600	200	...
3	2	CO	5	5	...	676524	231529		4500	100	...
3	3	SFD	1	4	...	676987	237225		600	50	...
3	4	SFD	3	4	...	676321	239445		1100	140	...
3	5	SFD	8	6	...	678463	238542		0	234	...

(Source: Developed for this research)

4.6 Variables for PF REP Characterisation

Geltner & Miller (2001) argue that RE characteristics such as property usage, location, building age and building size are expected to be the most important variables that characterise a RE. The variables available for this study include these most important physical parameters for the PF REP characterisation. The seventeen independent variables that are collected for each building included in the REs-ZH-Dataset are described below and summarized in Appendix 8.1.

Structural variables

1- Usage — contains the usage in percent of the building type between single family dwelling (SFD) and condominium (CO). It is calculated by dividing the number of SFD with the number of CO that are present into the PF's REP.

2- Age — contains the age of the building in years and it is calculated by subtracting the building's construction year from the year 2007 (the dataset was generated on August 13, 2007).

3- NrRooms — contains the number of rooms for the building.

4- Surface — contains the usable surface in square metres for the building.

5- Volume — contains the usable volume in cubic metres of the GVZ-Object. According to the GVZ (2008) definition, this volume includes the REs that are situated in the same physical building.

Macro Variables

6- GeoLocationX — indicates the location of the building with the geographical coordinate X in metres.

7- GeoLocationY — indicates the location of the building with the geographical coordinate Y in metres.

8- DistanceToCentre — contains the shortest distance in metres to a city (Zurich or Winterthur) that is considered as a regional centre for its attractiveness and its size.

9- TaxationLevel — contains the taxation level in percent of the domicile in which the building is situated.

Micro Variables

10- Lake	contains the visible hectares of lake seen from the building.
11- View	contains the visible hectares of surface (without lake) seen from the building.
12- Slope	contains the ground inclination in percent where the building is situated.
13- SunJuly	contains the mean of the sunshine power in Kilo Joule per square metres for the month July 07.
14- Dist_Bus	contains the distance in metres to the nearest bus station
15- Dist_Railway	contains the distance in metres to the nearest rail way station
16- Dist_School	contains the distance in metres to the nearest school
17- Walk_Index	contains the walk index, which has a range starting from zero up to 100. This index measures the on-foot attainability of points of interest such as food stores, pharmacies, restaurants, doctors, banks, bars, etc. For the calculation of this index, a distance less than 500 metres is considered as optimum, a distance between 500 metres and 1,250 metres as sub-optimum and distances over 1,250 metres are not considered relevant. The index is calculated by the STA and it is public available (Appendix 8.2).

4.7 Data Analysis Processing

The quantitative analysis of the data is undertaken using the computer software Statistical Package for Social Sciences (SPSS) and includes the following steps:

1. An explorative analysis to screen the data.
2. A descriptive analysis in order to gain a better understanding of the data. This includes a graphical examination, a check of the dataset for errors, missing data and outliers and the analysis of frequency distributions to gain an overall view of the raw data.
3. The normalisation of the measurements in order to reach unit-less variables to be used in the computation of the REP_benchmark.
4. The determination of the "distance" between two RE portfolios using the Euclidean distance measure.
5. Development of the REP Empirical Decision Model (REP-EDM) using a multivariate regression analysis that includes different statistical tests considered as scientifically appropriate and used in such a manner as to ensure validity and reliability of the developed model. While testing the residuals, normality is assumed for the distribution. If this is not the case, transformations are undertaken to deal with problems such as skewness and kurtosis (Hair et al. 2006).

4.8 Computation of the Pension Funds REP_benchmark

The RE physical characteristics are measured using different units. This increases the difficulty of comparison and consolidation or aggregation of the variables into a benchmark. In order to build a benchmark that can be used as a reference to compare the various RE characteristics, it is necessary to transform the measurements so that they are independent from the original units. This can be reached by converting absolute measurements into relative values such as (1) calculating a coefficient of variation; (2) calculating a percentage; or (3) normalising the measurement with a reference (Levine et al. 2005). In the following subsections, these three normalisation methods are described. Following this, the REP_benchmark is presented.

4.8.1 Normalisation - Coefficient of Variation

According to Levine et al. (2005), the coefficient of variation is a normalised measure of the dispersion of data points in a data series around the mean. It is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. The coefficient of variation (cv) represents the ratio of the standard deviation (σ) to the mean (μ) and it is calculated using Formula 4.1.

$$cv = \frac{\sigma}{\mu} \tag{4.1}$$

The coefficient of variation (cv) is only defined for non-zero mean and is most useful for variables that are always positive. It is useful because the standard deviation of data have always to be understood in the context of the mean of the data. The coefficient of variation is a dimensionless and scale invariant number. So when comparing between datasets with different units or very markedly different means, it is appropriate to use the coefficient of variation for comparison instead of the standard deviation (Levine et al. 2005).

A PF's REP includes several REs that are characterised by micro-, macro- and structural-variables as described in Appendix 8.1. Within the PF's REP, the coefficients of variation are calculated for each of the non-categorical RE's variables using the Formula 4.2.

$$cv_{ij} = \frac{\sigma_{ij}}{\mu_{ij}} \tag{4.2}$$

Where

cv_{ij}	: coefficient of variation for variable *i* and PF *j*
σ_{ij}	: standard deviation for variable *i* and PF *j*
μ_{ij}	: mean for variable *i* and PF *j*
i	: variable for RE characteristic
j	: pension fund

The next step, in order to build the multidimensional pension fund REP_benchmark (REP_B) that includes the entire PF population of the Canton Zurich, is to calculate a reference that indicates the central tendency for each individual variable over all PFs analysed. Kachigan (1991) defines central tendency as a single summary value that suggests a representative observation at which value the observations tend to be centred or to be clustered. While such a measure sacrifices much of the information inherent in the frequency distribution, it nonetheless serves as a very concise description of a body of data. He lists the most useful measures of central tendency: Mode, median, arithmetic mean, geometric mean, weighted mean, mean of grouped data and quartiles.

According to Kachigan (1991) the most important measure of central tendency and one of the basic building blocks of all statistical analyses is the arithmetic mean. The major advantage of the arithmetic mean over the mode and the median is that it takes into account the numerical value of every single observation in the data distribution and it represents a balance point or centre of gravity in that the sum of the distances to the observations below it is equal to the sum of the distances to the observations above it. For the non-categorical RE variables, the coefficient of variation used as a benchmark (unit-less) of a specific RE variable, is calculated with the arithmetic mean using Formula 4.3.

$$cv_{iB} = \frac{\sum_{j=1}^{n} cv_{ij}}{n} \qquad (4.3)$$

Where

cv_{iB}	: benchmark of coefficient of variation for variable *i*
cv_{ij}	: coefficient of variation for variable *i* and PF *j*
i	: variable for RE characteristic
j	: pension fund
n	: number of pension funds in Canton Zurich

4.8.2 Normalisation - Percentage

For the categorical variable "Usage," the normalisation is undertaken using the percentage computation of the proportion between the number of single family dwelling (SFD) and the number of condominiums (CO) into the PF's REP as showed in the Formula 4.4.

$$Usage_j = \frac{Number\_SFD_j}{Number\_CO_j} * 100 \tag{4.4}$$

Where

$Usage_j$: variable “Usage” in percent for PF j

j : pension fund

The coefficient of variation for the variable “Usage” as benchmark (unit-less) for all the PFs in Canton Zurich is calculated directly, starting from the percentage values using Formula 4.5.

$$cv_{1B} = \frac{\sigma_1}{\mu_1} \tag{4.5}$$

Where

cv_{1B} : benchmark of coefficient of variation for variable “Usage”

σ_1 : standard deviation for RE variable 1 (“Usage”) for all PFs

μ_1 : mean for RE variable Nr. 1 (“Usage”) for all PFs

4.8.3 Normalisation - Reference

The arithmetic mean of a specific RE variable of a PF is used to determine the mean between all PFs and is calculated using the Formula 4.6.

$$\mu_{iB} = \frac{\sum_{j=1}^{n} \mu_{ij}}{n} \tag{4.6}$$

Where

μ_{iB} : benchmark of mean between all PFs for variable i

μ_{ij}	: mean for variable *i* and PF *j*
i	: variable for RE characteristic
j	: pension fund
n	: number of pension funds in Canton Zurich

The benchmark of the mean between all PFs for the specific variable *i* (μ_{iB}) is used as reference for the linear normalisation. Thus, the linear normalisation is applied between the mean of each PF's variable (μ_{ij}) and the mean of all PFs' means for the same variable (μ_{iB}) using Formula 4.7, where the minimum value is transformed to zero and the mean of all PFs is set to one.

$$NormV_{ij} = \frac{\mu_{ij}}{\mu_{iB}} \tag{4.7}$$

Where

$NormV_{ij}$	: linear normalised value of mean for variable *i* and PF *j*
μ_{iB}	: benchmark of mean between all PFs for variable *i*
μ_{ij}	: mean for variable *i* and PF *j*
i	: variable for RE characteristic
j	: pension fund

The result is a unit-less variable ($NormV_{ij}$) containing the linear normalised value of the mean of each PF's variable (μ_{ij}). The method used for the linear normalisation is presented in Appendix 8.3.

4.8.4 Multidimensional PFs REP_benchmark

The multidimensional pension fund REP_benchmark (REP_B) can be represented by Formula 4.8 as a function of the normalised variables ($NormV_{iB}$), the coefficients of variation (cv_{iB}) and the means of the variables for each PF (μ_{iB}).

$$REP_B = f(NormV_{1B},...,NormV_{iB},...,NormV_{mB},$$
$$cv_{1B},..cv_{iB},...,cv_{mB},\mu_{1B},..\mu_{iB},..\mu_{mB}) \qquad (4.8)$$

Where

REP_B	: REP_benchmark of the PF RE market
$NormV_{iB}$	: benchmark as normalised mean of variable *i*
cv_{iB}	: benchmark of coefficient of variation for variable *i*
μ_{iB}	: mean between all PFs for variable *i*
i	: variable for a RE characteristic
m	: number of variables for RE characteristics

The normalised variables are used to develop the REP-EDM. The coefficients of variation and the means of the variables for each PF are used directly in the REP analysis. The former are used as a relative measure of the scatter in the data relative to the mean of the variable and the latter are used as an indication of the absolute values of the RE market.

4.9 "Distance" between a PF's REP and the PFs' REP_benchmark

The statistical procedure for the measurement of the "distance" is determined through an analysis of various possible measures of similarity and distance between a pair of objects. According to Kachigan (1991) the most commonly used measures of similarity and distance are: a) Correlation coefficients that reflect the similarity between a pair of objects measured on several variables; b) Euclidean distance based on the Pythagorean Theorem; c) matching-type measure of similarity used when objects are measured on dichotomous variables; and d) direct scaling of similarities including expert judges that rate the similarity of each object to each of the other objects.

Because in this study a ranking among different RE portfolios has to be built, no binary variables are included in the calculation and no human judgement is desired in order to assure objectivity of the model. This being the case, the Euclidean Distance (ED) appears to be the most appropriate measure of similarity to calculate the multivariate distance between the PFs REP_benchmark (REP_B) and a specific PF's REP (REP_S). The calculation is given by Formula 4.9.

$$ED_j = \left[\sum_{i=1}^{m} (NormV_{iB} - NormV_{ij})^2 \right]^{1/2} \tag{4.9}$$

Where

ED_j	: Euclidean Distance of PF *j* to benchmark
$NormV_{iB}$	: benchmark for normalised mean of variable *i*
$NormV_{ij}$	: linear normalised value of mean for variable *i* and PF *j*
i	: variable for RE characteristic
j	: pension fund
m	: number of variables for RE characteristics

In this study, the Euclidean distance represents a measure of the divergence of a PF's REP and the REP_benchmark. As explained earlier, this may be viewed as an indicator of liquidity risk (idiosyncratic risk) of the pension fund under analysis and furnishes a ranking among the various PFs of the Canton Zurich. The larger the "distance" to the benchmark the larger the potential trading-related liquidity risk and therefore the lower the ranking will be.

4.10 Random Selection of a PF for Discussion and Practical Analysis

Before beginning the development of the REP-EDM using a multivariate regression analysis based on the PF samples, a single PF is selected randomly and extracted from the dataset. The goal is to develop a regression model without this selected PF and, later in the Chapter Six, this PF is analysed with the model built independently of it. The results achieved by introducing the extracted PF into the model are then discussed and analysed.

4.11 Development of the REP Empirical Decision Model (REP-EDM)

Hair et al. (2006, p. 13) affirm that multivariate regression analysis can be used to identify and analyse the relationship between the independent and dependent variables. In order to choose the multivariate techniques to use in a study, the researcher must answer three basic questions about the research objective and the nature of the data to analyse. The first question is 'can the variables be divided into independent and dependent in a single analysis?' and the answer for this study is "yes." The second question is 'how many variables are treated as dependent in a single analysis?' and the answer for this study is "one dependent variable in a single relationship." The third question is 'how are the variables, both dependent and independent, measured?' and the answer for this study is "metric for both." The selected statistical analysis methodology for this study is multiple-regression.

Equating the PF's rankings, represented by the Euclidean distance (ED), with the means of RE physical characteristics for each PF using a multivariate regression in a iterative process, it is possible to create a model for the evaluation of a specific PF's REP in comparison to the PF REP_benchmark. The **REP-EDM** is shown in Formula 4.10.

$$ED_j = \beta_{0j} + \sum_{i=1}^{m} \beta_{ij} \bar{x}_{ij} + \varepsilon_j \qquad (4.10)$$

Where

ED_j	: Euclidean distance of PF *j* to benchmark
β_i	: regression's coefficient for variable *i* and PF *j*
$\bar{x}_{ij}$	: mean for variable (RE characteristic) *i* and PF *j*
i	: variable for RE characteristic
m	: number of variables for RE characteristics
j	: pension fund
n	: number of pension funds in Canton Zurich
ε_j	: error term for PF *j*

The objectives for using the regression analysis in this study are: 1) To determine whether or not a relationship exists between the variables; 2) to describe the nature of the relationship; 3) to assess the degree of accuracy of description or prediction achieved by the regression equation; 4) to assess the relative importance of the various predictor variables in their contribution to variation in the criterion variable; and 5) to validate the developed model through statistical analysis. A significant overall relationship between the dependent variable and the independent variables in the developed model can be confirmed when the statistical ANOVA F-test rejects the following hypothesis:

H_o : all regression coefficients are equal to zero

This regression model allows the evaluation of a specific PF's REP (REP_S) starting with its physical characteristics in comparison to the PF REP_benchmark (REP_B). The information about the "distance" and the ranking compared to the other PFs in Canton Zurich can be used to set the baseline for the PF's decisions for an optimisation of the investments in the REP with the possibility to replicate the RE market, reducing the divergence between the PF's REP and the benchmark and, potentially, the liquidity risk (non-systematic risk) of the PF.

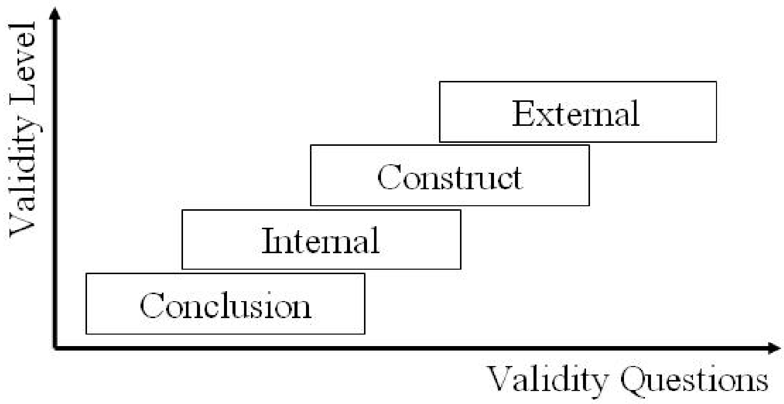
4.12 Quality of Research - Validity and Reliability

Leedy and Ormrod (2005) and Neuman (2003a) state that the scientific community is governed by norms and values that serve as guidelines for all researchers and that are tested by certain criteria that must be built into the research design and methodology. They define the following four criteria: 1) Universality, in which irrespective of who conducts research and regardless of where it was conducted, the research is to be judged only on the basis of scientific merit; 2) replication, which focuses on the repeatability of a research; 3) control, in which the researcher must isolate those factors that are central to the research problem or, control phenomena and factors that are not under study, important for replication and for consistency issues and lastly; 4) measurement, in which data should be able to be measured in some way. Validity and reliability are essential in order to fulfil these four criteria. Thus in the following subsections, validity and reliability are assessed under consideration of these criteria.

4.12.1 Validity Overview

Validity concerns whether the concepts being investigated are actually the ones being measured or tested. It must always be verified independently from the level of constraint and must be understood in relative terms because the absolute validity cannot be achieved (Graziano & Raulin 2000). Trochim (2006, p.1) defines validity as '…the best available approximation to the truth of a given proposition, inference or conclusion.' Trochim (2006) and Yin (1994) subdivide the validity into four types: Conclusion, internal, construct and external (Figure 4.4). In the following subsections, these four validity types and the reliability are discussed for this study.

Figure 4.4 Types of Validity – Cumulative Questions



(Source: Adapted from Trochim 2006)

4.12.2 Conclusion Validity

Conclusion validity, also called statistical validity, '... relates to the use of appropriate analytical tools and the power, or ability to detect a relationship, of the statistics used'

(USQ 2006, p. 2.14). The question addressed by this type of validity is ‘is there a relationship between the variables?’ (Trochim 2006, p. 1). When researchers use statistical procedures to test the null hypothesis, they are making a statement about statistical validity of the results, testing if the results are due to some systematic factor or if they are due merely to chance variation. Rejecting the null hypothesis is a necessary first step in testing the effects of the independent variable (Graziano & Raulin 2000). In this study, the conclusion validity is verified by testing the hypothesis about the significance of the regression coefficients in the multiple regression of the developed REP-EDM.

4.12.3 Internal Validity

Internal validity of a research study is the extent to which its design and the data it yields allow the researcher to draw accurate conclusions about cause-and-effect and other relationships. The question addressed by this type of validity is ‘is the relationship causal?’ (Trochim 2006, p. 1). A problem that increases difficulty to draw casual inference between the independent variables and the dependent variable may be due to some extraneous variables that influence the research. According to Graziano and Raulin (2000), the best time to rule out confounding variables is during the design phase, when the researchers anticipate possible confounding variables and design controls to eliminate their effect on the dependent variable.

This study is a descriptive analysis of the PFs REP in Canton Zurich and the principal focus is not to find out the causes and reasons of the current RE market situation, but rather to evaluate a PF's REP basing the analysis on existing data by finding a relationship respectively, correlation between the PF's REP physical characteristics and the REP_benchmark. Having determined this correlation, a meaning, a predictive function and an interpretation can be extracted from the data. Attention must be paid to the interpretation because when variables are correlated, this does not necessary mean that one variable can influence the others or vice versa. In fact, a cause-and-effect relationship on the basis of a correlation alone can never be inferred (Leedy & Ormrod 2005). Finally, this correlational and descriptive research is based on an existing physical RE dataset which contains quantifiable and standardised measurements that are objective (not perceived) dimensions based on physical attributes. This enhances the internal validity allowing an interpretation of the results.

4.12.4 Construct Validity

The construct validity is the '…approximate truth of the conclusion that your operationalisation accurately reflects its construct' (Trochim 2006, p. 1). In this study, the correct operational measures for the concepts being studied are enhanced by the Swiss law that imposes standardised measurements for the RE that have to be reported into the balance sheet of the PF. The data is measured with RE physical characteristics and no human interpretation is done, thus a maximum of objectivity in the measurements can be reached. Additionally, the quality of the data is assured because it is used for official purpose by STA.

The construct validity of this research is also enhanced by the fact that the results are reached systematically, guaranteeing reproducibility, universality and confidence in the results. An example is the fact that to compare different pension funds (PFs), it is necessary to have the same reference. In other words, a comparison can be made only between pension funds which have the same legal basis. This is assured by registered PFs and this study analyses exactly this type of PF in Canton Zurich. Additionally, according to Leedy and Ormrod (2005), this validity can be enhanced using primary data and not secondary data and, in this study, only primary data is treated, increasing validity.

4.12.5 External Validity

'The external validity of a research study is the extent to which the conclusion drawn can be generalised to other contexts' (Leedy & Ormrod 2005, p. 99). In order to assess whether this study has external validity, it is necessary to ascertain whether the findings apply to PF whose place, times, and circumstances differ from those of study PFs. A study that randomly selects PF from the most diverse and representative populations is more likely to have external validity than one that does not. The question addressed by this type of validity is 'can we generalise to other persons, places and times?' (Trochim 2006, p. 1).

This study considers not only samples of the PFs analysed, but it considers the entire population of PFs in Canton Zurich. Therefore, a generalisation within the defined geographical limitation seems to be possible. Although Hair et al. (2006, p. 9) state that '…a census of the entire population makes statistical inference unnecessary, because any difference or relationship however small, is "true" and does exist,' the external validity in this study cannot be confirmed. The geographical limitation and the fact that RE is a local market reduce the interpretation of the results only to the Canton Zurich. Although the results cannot be generalised to be applied in a larger context such as Switzerland, the developed model REP-EDM can be used in a general way and in larger regions. This is due to the transparency in the standardised RE measurements used and in the statistical analysis done, which assure replication in other areas.

4.12.6 Reliability

Hair et al. (2006, p.8) define reliability as '…the degree to which the observed variable measures the "true" value and is error free.' 'Reliability is a necessary but insufficient condition for validity' (Leedy & Ormrod 2005, p. 29). The validity and reliability of a measurement instrument influence the extent to which something can be learned about the phenomenon under study, the probability to obtain statistical significance in the data analysis and the extent to which meaningful conclusions can be drawn from the data. The validity of a measurement instrument is the extent to which the instrument measures what it is supposed to measure and the reliability is the consistency with which a measuring instrument yields a certain result when the entity being measured has not changed (Leedy & Ormrod 2005).

In this study, reliability is enhanced by Swiss law which imposes standardised measurements for the RE physical characteristics. The raw data received from the statistical office of the Canton Zurich (STA), the building insurance Zurich (GVZ), the

Swiss federal government for statistics (SFG) and the Zurich cantonal bank (ZKB) has the necessary quality to guarantee validity and reliability. In fact, the primary data collection done by these sources is regulated by Swiss law, which defines the level of quality needed to guarantee validity and reliability of the measurement system. STA uses this data also for official government statistics, therefore the data is of high quality and reliability. Another important factor for the reliability is that the model must be independent from the units used by the measurement of the variables. This has been reached by converting absolute measurements into relative values calculating coefficients of variation, percentages and normalised values.

Reliability is also understood as the accurate and precise reproducibility of the measurements with the same results (Kachigan 1991) and this is the case in this study. In fact, the data used in this study comes from standardised measurable RE physical characteristics, and is, therefore, not influenced by human opinions or bias. In addition, the same dataset used in this research could be obtained from other regions with the same quality and the major problem is not the reliability of the measurements, but rather to obtain the data and permission to carry out the analysis.

4.13 Limitations

A limitation of this study is the correlational research methodology itself, which has some inherent limitations. Correlational methodology is a non-experimental research method and according to Kerlinger and Lee (2000) it has three major weaknesses: 1) The inability to manipulate independent variables; 2) the lack of power to randomise; and 3) the risk of improper interpretation. Compared to experimental research, other things being equal, non-experimental or descriptive research lacks controls, because it is based on existing data coming from the past and that has not been explicitly collected for this study.

Because it is not possible to manipulate the independent variables, the analysis and interpretation of the results derived from this dataset need to be undertaken carefully considering the possible confounding variables, the level of constraint imposed by the method and the impossibility to infer a cause-and-effect relationship on the basis of correlation alone. Leedy (2005, p.182) confirms that '…simply put, correlation does not, in and of itself, indicate causation.' Graziano and Raulin (2000, p. 151) affirm that '... although a consistent correlation does not imply causality,' it can be used to predict future events. Kachigan (1991, p. 134) adds that '…the existence of a correlation between two variables does not imply causality' (Levine et. al 2005; Zikmund 2003).

The seventeen independent variables available for this study that are collected for each building included in the REs-ZH-Dataset are described in the Section 4.6 and summarized in Appendix 8.1. A limitation of this study is that not all existing RE characteristics have been included into the model. Although the most relevant RE variables are available and have been obtained from GVZ and STA, these do not include all possible variables that can be considered to describe a RE. In fact, independent variables not considered which could increase the precision of the result are, for example, the condition of the building, quality of the location, aesthetically dimension and type of municipality, economic factors such as the current global and local markets, legal constraints, buyer or seller characteristics and demography of population.

4.14 Conclusion

This chapter presented a justification for the positivist research paradigm used for this study. It also presented a review and defence of the chosen research design as descriptive cross-sectional analysis of quantitative data and the use of the correlational research methodology. The correlational research methodology is suitable for this study for three reasons. First, this research involves metric data. Second, no control group can be used, because the data is historical. Third, the aim of the relationship analysis between a PF's REP and the REP_benchmark is not to find out the causes and reasons of the current RE market situation, but rather to evaluate a PF's REP compared to a REP_benchmark basing the analysis on existing data.

The sampling design has been defined including the entire PF population of the Canton Zurich and the data collection and consolidation procedures have been outlined. The statistical analysis of the data has been chosen and the development of the REP-EDM described using a multivariate regression analysis that includes different statistical tests used to ensure validity and reliability of the developed model. The developed REP-EDM model allows the researcher to evaluate a specific PF's REP in comparison to the PF REP_benchmark, contributing to increase the understanding, the body of knowledge and the transparency of the RE market. Finally, the limitations of the methodology used have been identified and the ethical issues have been addressed.

5 ANALYSIS OF RESULTS

5.1 Introduction

The preceding chapters identified the research issues, outlined the importance of this study through a literature review, defined the research paradigm, presented the research design, the research methodology used for this research as well as this study's limitations. This chapter reports and presents results derived from the quantitative data analysis and focuses on the development of the REP-EDM following the procedure described in the previous research methodology chapter.

In Section 5.2, the collected and consolidated real estate dataset of the Canton Zurich (*REs-ZH-Dataset*) is described. Section 5.3 presents the data processing procedures such as screening, editing and transforming the data and develops a descriptive analysis in order to better understand the *REs-ZH-Dataset*. Section 5.4 considers the computation of the PF REP_benchmark and Section 5.5 depicts the calculation of the multivariate "distance" between the PFs REP_benchmark (REP_B) and a specific PF's REP (REP_S). The result is saved into a sub dataset called *PFs-ZH-Dataset* which contains: a) Only the selected PFs of the Canton Zurich owning REs, b) the respective PF Euclidean distance, and c) the calculated means of their RE variables. In Section 5.6, the PFs-ZH-Dataset is undertaken to a further data processing analysis. In Section 5.7, the REP-EDM is built using a multivariate regression model. In Section 5.8 conclusions are drawn.

As proposed by Coolican (1990), the statistical tests to validate the developed REP-EDM model are reported in the following three levels and are based on “p” (the probability). The first level is defined as “significant” when the range of the p-value is: $0.05 > p < 0.01$. The second level is defined as “highly significant” when the range of the p-value is: $0.01 > p < 0.001$. The third level is defined as “very highly significant” when the range of the p-value is: $0.001 > p$. When the p-value does not fall below 0.05, the finding is considered not significant. The reported probabilities are based on two-tailed tests as the comparisons have two possible directions.

5.2 Data Collection and Consolidation

On the August 13, 2007, GVZ delivered to STA 6,008 names and addresses of the building owners in the Canton Zurich excluding private owners, because these are not relevant for a study about PFs. The dataset included 368,011 REs owned by these 6,008 legal entities. At the same time, the SFG has delivered to STA 728 PF names representing the entire PF population with domicile in the Canton Zurich. The match of the PF names with the building owners results in 76 PFs that own 17,126 REs in Canton Zurich. The dataset is saved by the STA in a database called REs-ZH-Dataset. After a geo-codification of the addresses with the ZKB GIS information system it is possible to add the 12 micro and macro-characteristics to each of the 17,126 REs. Finally the five building structural characteristics from the STA are added to each record to complete the REs-ZH-Dataset that forms the basis dataset for this study. In brief, the REs-ZH-Dataset includes for a cross-sectional analysis at the August 13, 2007, 76 PF owners with their 17,126 REs and their 17 variables described in Appendix 8.1. The Table 5.1 summarises the steps taken to construct the REs-ZH-Dataset.

Table 5.1 *REs-ZH-Dataset* Construction Process

Data Collection and Consolidation		Sample N
(1) GVZ data, 13.08.07		
	Number of RE owners	6'008
	Number of REs	368'011
(2) SFG data, 13.08.07		
	Number of PFs in Canton Zurich	728
(3) GVZ data (1) matched with SFG data (2) resulting in *REs-ZH-Dataset*		
	Number of PF owners	76
	Number of REs	17'216
(4) ZKB data with GIS System added		
	Macro- and Microvariables	12
	Number of REs	17'216
	Total added fields	206'592
(5) STA data with GIS System added		
	Structural variables	5
	Number of REs	17'216
	Total added fields	86'080
REs-ZH-Dataset		
-> original full sample size, before analysis of missing data, errors and outliers		
	Number of PF owners	76
	Number of REs	17'216
	Structural, Macro- and Microvariables	17
	Total number of fields	292'672

(Source: Developed from field data)

5.3 *REs-ZH-Dataset* - Data Screening and Transformations

Data screening (parsing) is used in this study to make sure that data have been entered correctly. It includes exploring and selecting raw data, checking data for missing values, errors and outliers. According to Levine et al. (2005), the three major properties used to examine and understand data are the measures of location or central tendency, the spread or variability and the shape. Leedy and Ormrod (2005) add the correlation or the extent to which different variables are related to one another. These properties set the basis for the analysis of the data.

5.3.1 Missing Values Analysis

The 17 variables described in Appendix 8.1 have been analysed for missing values. The result is depicted in Table 5.2 and shows that three variables "Lake_ha", "View_ha" and "Slope_Percent" contain two, eight and two missing values respectively. According to Hair et al. (2006) and Coakes and Steed (2007) finding a remedy for missing data can be a practical solution such as deleting the cases or variables, using a mean substitution or calculating a linear interpolation from adjacent points. A main issue is to find the appropriate remedy for the missing value that does not introduce possible bias or undesirable effects in the variable. Through data analysis, the remedies used for the imputation of the missing values are described below.

Table 5.2 Case Processing Summary – Missing Values

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Usage	17126	100.0%	0	.0%	17126	100.0%
Age	17126	100.0%	0	.0%	17126	100.0%
NrRooms	17126	100.0%	0	.0%	17126	100.0%
Surface_m2	17126	100.0%	0	.0%	17126	100.0%
Volume_m3	17126	100.0%	0	.0%	17126	100.0%
GeoLocationX	17126	100.0%	0	.0%	17126	100.0%
GeoLocationY	17126	100.0%	0	.0%	17126	100.0%
DistanceToCentre	17126	100.0%	0	.0%	17126	100.0%
TaxationLevel	17126	100.0%	0	.0%	17126	100.0%
Lake_ha	17124	100.0%	**2**	.0%	17126	100.0%
View_ha	17118	100.0%	**8**	.0%	17126	100.0%
Slope_Percent	17124	100.0%	**2**	.0%	17126	100.0%
SunJuly_KJ_m2	17126	100.0%	0	.0%	17126	100.0%
Dist_Bus_m	17126	100.0%	0	.0%	17126	100.0%
Dist_Railway_m	17126	100.0%	0	.0%	17126	100.0%
Dist_School_m	17126	100.0%	0	.0%	17126	100.0%
Walk_Index	17126	100.0%	0	.0%	17126	100.0%

(Source: Developed from field data using SPSS)

The variables “Lake_ha” and “Slope_Percent” have two missing values that have been imputed using the mean of the adjacent geo-coordinate values with the assumption that near (adjacent) REs have similar physical characteristics such as view of the lake and slope. The missing values of the variable “View_ha” have been imputed using a linear interpolation (Appendix 8.4). The assumption of linear interpolation is justified by the topography as precise description of a place and by the near distance with the adjacent geo-coordinates. The analysis of the rest of the variables demonstrates that the REs-ZH-Dataset does not contain any other missing values.

5.3.2 Errors and Outliers Analysis

The 17 variables described in Appendix 8.1 have been analysed for errors and outliers and the result is a reduction of the REs-ZH-Dataset samples from 76 PFs to 74 PFs and from 17,126 REs to 15,836 REs. The difference of 1,290 REs (samples) that represents a reduction of seven percent of the original sample size consists of incomplete samples that contain errors or outliers, therefore they have been eliminated from the analysis. According to Studenmund (2006, p. 401) '…if a few observations have incomplete data in a cross-sectional study, they can be dropped from the sample.' The details about the REs-ZH-Dataset analysis are presented in the following.

All of the 17 variables of the REs-ZH-Dataset are tested with the same procedure that includes a descriptive and an explorative analysis. According to Hair et al. (2006), observations that contain errors are often problematic, may not be representative of the population, may be counter to the objectives of the analysis and can seriously distort statistical tests. Therefore, they should be corrected or eliminated. In contrast, outliers cannot be categorised automatically as beneficial or problematic without having considered the context of the analysis and the types of information that they may provide.

Outliers are analysed with extreme values statistics and with the box-plot that summarises information about the distribution of scores. In fact, it plots summary statistics such as the median, 25^{th} and 75^{th} percentiles and extreme scores in the distribution (Coakes & Steed 2007). The three variables "NrRooms", "Surface" and "Volume" that contain samples with errors and outliers are analysed and discussed in depth. The summarised analyses of the variables that do not indicate evident inconsistency are available in Appendix 8.5/8.6.

I) Analysis of the variable "NrRooms"

Table 5.3 indicates a mean value of 3.19 rooms per RE and a standard deviation of 1.16 rooms. The tests of Skewness (3.31) and Kurtosis (92.27) do demonstrate evidence of a violation of the normality assumption. According to Garson (2007), their values should be within the +2 to −2 range when the data is normally distributed. Another indication of non-normality is confirmed by the outliers present in the Figure 5.1. The first problem can be identified by the minimum values of zero and it indicates a possible error in the data entry. It is not plausible that a RE has zero rooms, therefore the 59 samples having a zero value has been deleted. From the Tables 5.4 and the Figure 5.1 additional problems are identified. In fact, seven samples indicate a number of rooms per RE greater than nine with a maximum value of 41, thus these seven samples has been deleted from the REs-ZH-Dataset.

Table 5.3 Descriptive Statistic for Variable "NrRooms", N=17,126

			Statistic	Std. Error
NrRooms	Mean		3.1906	.00891
	95% Confidence Interval for Mean	Lower Bound	3.1732	
		Upper Bound	3.2081	
	5% Trimmed Mean		3.1946	
	Median		3.0000	
	Variance		1.359	
	Std. Deviation		1.16570	
	Minimum		.00	
	Maximum		41.0	
	Skewness		3.314	.019
	Kurtosis		92.275	.037

(Source: Developed from field data using SPSS)

Table 5.4 Explorative Statistic for Variable “NrRooms”, N=17,126

a) Case Processing Summary

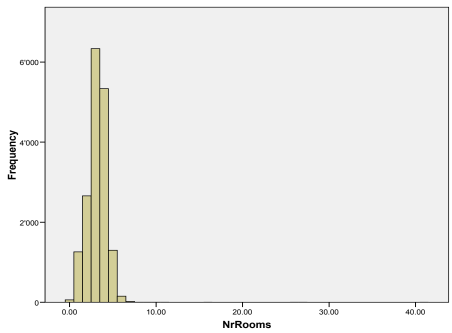
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
NrRooms	17126	100.0%	0	.0%	17126	100.0%

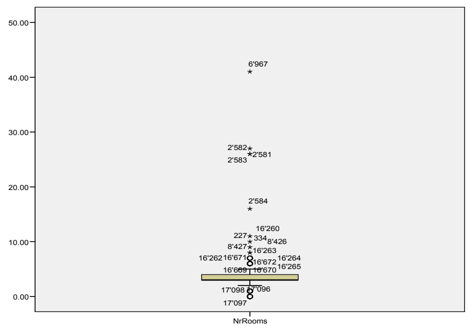
b) Extreme Values

			Case Number	Value
NrRooms	Highest	1	6967	41.0
		2	2581	27.0
		3	2582	26.0
		4	2583	26.0
	Lowest	1	17099	.00
		2	17098	.00
		3	17097	.00
		4	17095	.00

(Source: Developed from field data using SPSS)

Figure 5.1 Histogram and Boxplot for Variable “NrRooms”, N=17,126





(Source: Developed from field data using SPSS)

Following the deletion of the questionable 66 samples (N=17,060), no further errors or problems remain in need of special treatment (Appendix 8.25). Following the deletion of the 66 samples, the mean remained practically at the same level of 3.19 rooms per RE and the standard deviation has decreased by 9.5 percent to 1.06 rooms. The Kurtosis and

Skewness values indicate an approximate normal distribution of the data. The resulting REs-ZH-Dataset after the corrections on the variable "NrRooms" contains one PF and 66 REs less than in Table 5.1 and is depicted in the Table 5.5.

Table 5.5 *REs-ZH-Dataset* after Corrections on the Variable "NrRooms"

REs-ZH-Dataset
-> sample size, after corrections on the variable "NrRooms" (-66)

Number of PF owners	75
Number of REs	17'060
Structural, Macro- and Microvariables	17
Total number of fields	290'020

(Source: Developed from field data)

<u>II) Analysis of the variable "Surface"</u>

Table 5.6 indicates a mean value of 75.28 square metres surface per RE and a standard deviation of 31.28 square metres. The Kurtosis and Skewness values indicate an approximate normal distribution of the data.

Table 5.6 Descriptive Statistic for Variable "Surface", N=17,060

			Statistic	Std. Error
Surface	Mean		75.2838	.23951
	95% Confidence Interval for Mean	Lower Bound	74.8144	
		Upper Bound	75.7533	
	5% Trimmed Mean		75.8103	
	Variance		978.656	
	Std. Deviation		31.2834	
	Minimum		.00	
	Maximum		360.00	
	Skewness		-.054	.019
	Kurtosis		1.609	.038

(Source: Developed from field data using SPSS)

From the Table 5.6, 5.7 and the Figure 5.2, a problem can be identified by the minimum value of zero that indicates a possible error in the data entry. It is not plausible that a RE has zero square metres of surface area. Therefore the 845 samples having a zero value have been deleted from the REs-ZH-Dataset.

Table 5.7 Explorative Statistic for Variable "Surface", N=17,060

a) Case Processing Summary

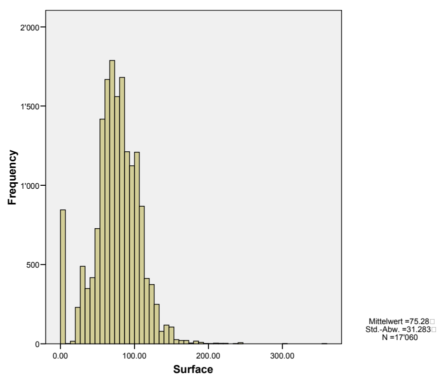
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Surface	17060	100.0%	0	.0%	17060	100.0%

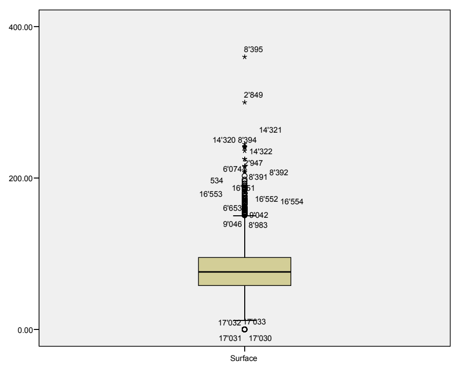
b) Extreme Values

			Case Number	Value
Surface	Highest	1	8395	360.00
		2	2849	300.00
		3	6072	245.00
		4	2779	242.00
	Lowest	1	17033	.00
		2	17032	.00
		3	17031	.00
		4	17029	.00

(Source: Developed from field data using SPSS)

Figure 5.2 Histogram and Boxplot for Variable "Surface", N=17,060





(Source: Developed from field data using SPSS)

After the 845 samples have been deleted (N=16,215) the tables and figures in Appendix 8.26 do not indicate evidence of other errors or outliers that need special treatment. The mean has increased by 5.2 percent to 79.20 square metres surface and the standard deviation has decreased by 14.3 percent to 26.81 square metres. The Kurtosis and Skewness values indicate an approximate normal distribution of the data. The resulting REs-ZH-Dataset after the corrections on the variable "Surface" contains one PF and 845 REs less than in Table 5.5 and is depicted in the Table 5.8.

Table 5.8 *REs-ZH-Dataset* after Corrections on the Variable "Surface"

REs-ZH-Dataset		
-> sample size, after corrections on the variable "Surface" (-845)		
	Number of PF owners	74
	Number of REs	16'215
	Structural, Macro- and Microvariables	17
	Total number of fields	275'655

(Source: Developed from field data)

<u>III) Analysis of the variable "Volume"</u>

The Table 5.9 indicates a mean value of 9,303.83 cubic metres volume per GVZ-Object and a standard deviation of 11,205.92 cubic metres. The tests of Skewness (3.882) and Kurtosis (20.596) do demonstrate evidence of a violation of the normality assumption. From the Table 5.10 and the Figure 5.3 a problem can be identified by the minimum value of zero that indicates a possible error in the data entry. It is not plausible that a GVZ-Object has zero cubic metres of volume, therefore the 359 samples having a zero value have been deleted from the REs-ZH-Dataset.

Table 5.9 Descriptive Statistic for Variable "Volume", N=16,215

			Statistic	Std. Error
Volume	Mean		9303.830	88.001
	95% Confidence Interval for Mean	Lower Bound	9131.3384	
		Upper Bound	9476.3230	
	5% Trimmed Mean		7548.0339	
	Median		6074.0000	
	Variance		125572769.64	
	Std. Deviation		11205.92565	
	Minimum		.00	
	Maximum		160832	
	Range		160832.00	
	Interquartile Range		7255.00	
	Skewness		3.882	.019
	Kurtosis		20.596	.038

(Source: Developed from field data using SPSS)

Table 5.10 Explorative Statistic for Variable "Volume", N=16,215

a) Case Processing Summary

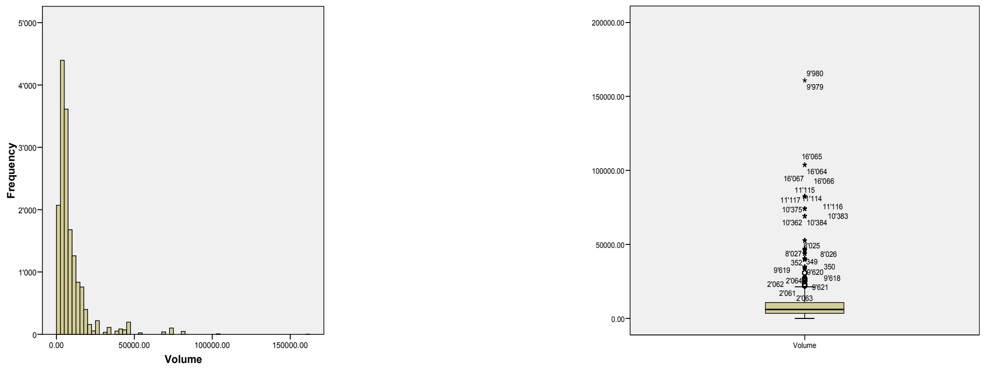
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Volume	16215	100.0%	0	.0%	16215	100.0%

b) Extreme Values

			Case Number	Value
Volume	Highest	1	9979	160832
		2	9980	160832
		3	16061	103779
		4	16062	103779
	Lowest	1	16126	.00
		2	16125	.00
		3	16124	.00
		4	16122	.00

(Source: Developed from field data using SPSS)

Figure 5.3 Histogram and Boxplot for Variable "Volume", N=16,215



(Source: Developed from field data using SPSS)

A second problem in twenty of the REs in the sample is the implausible relationship between the volume and the total number of rooms for the REs included into the GVZ-Object. Therefore the twenty outliers have been eliminated. After that, the 379 samples have been deleted (N=15,836) the tables and figures in Appendix 8.27 do not indicate evidence of other errors or outliers that need special treatment. The mean has increased by 1.7 percent to 9,460.23 cubic metres volume and the standard deviation has decreased by 2.4 percent to 10,940.15 square metres. The Kurtosis and Skewness values indicate a non-normal distribution of the data. The resulting REs-ZH-Dataset after the corrections on the variable "Volume" maintained the 74 PFs and is depicted in the Table 5.11.

Table 5.11 *REs-ZH-Dataset* after Corrections on the Variable "Volume"

REs-ZH-Dataset		
-> sample size, after corrections on the variable "Volume" (-379)		
	Number of PF owners	74
	Number of REs	15'836
	Structural, Macro- and Microvariables	17
	Total number of fields	269'212

(Source: Developed from field data)

IV) Summary of Errors and Outliers Analysis

In conclusion, a total of seven percent or 1,290 samples have been deleted from the REs-ZH-Dataset that contained originally 17,126 samples. These 1,290 samples have been excluded from the analysis, the sample size reached 93 percent (15,836/17,126) and it is representative because no systematic error has been made with the elimination of the inconsistent elements of the data.

5.3.3 Data Transformations

The explorative statistics of the variables still indicate the presence of outliers and various non-normal distributions of the variables. This is demonstrated in Table 5.12 with application of the Kolmogorov-Smirnov (KS) test. In this case, the Shapiro-Wilk (SW) is not applied because the KS significance is not higher than 0.05. In fact, according to Coakes and Steed (2007), only if the significance of KS is greater than 0.05 and SW is not too far below from 1.0, then a normal distribution can be assumed. Also the tests of Skewness and Kurtosis do demonstrate evidence of a violation of the normality assumption by various variables.

Table 5.12 Test of Normality for Variables in the *REs-ZH-Dataset*, N=15,836

a) Kolmogorov-Smirnov

	Kolmogorov-Smirnov (a)		
	Statistic	df	Sig.
Usage	0.526	15836	0
Age	0.048	15836	0
NrRooms	0.199	15836	0
Surface_m2	0.047	15836	0
Volume_m3	0.221	15836	0
GeoLocationX_m	0.126	15836	0
GeoLocationY_m	0.07	15836	0
DistanceToCentre_m	0.116	15836	0
TaxationLevel_Percent	0.26	15836	0

	Kolmogorov-Smirnov (a)		
	Statistic	df	Sig.
Lake_ha	0.364	15836	0
View_ha	0.047	15836	0
Slope_Percent	0.149	15836	0
SunJuly_KJ_m2	0.143	15836	0
Dist_Bus_m	0.142	15836	0
Dist_Railway_m	0.089	15836	0
Dist_School_m	0.127	15836	0
Walk_Index	0.104	15836	0

a Lilliefors Significance Correction

b) Skewness and Kurtosis

	N	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error		Statistic
Usage	15836	1.9931	-11.874	.019	139.015	.039
Age	15836	35.5318	.456	.019	.252	.039
NrRooms	15836	3.1714	-.124	.019	-.053	.039
Surface_m2	15836	78.5902	.637	.019	2.412	.039
Volume_m3	15836	9460.2390	3.622	.019	16.266	.039
GeoLocationX_m	15836	686947.28	.364	.019	-.745	.039
GeoLocationY_m	15836	249879.76	-.093	.019	-.143	.039
DistanceToCentre_m	15836	7969.1658	1.089	.019	.698	.039
TaxationLevel_Percent	15836	123.4610	-1.284	.019	.714	.039
Lake_ha	15836	514.1845	2.663	.019	7.075	.039
View_ha	15836	14945.675	.426	.019	.356	.039
Slope_Percent	15836	3.0454	1.420	.019	1.876	.039
SunJuly_KJ_m2	15836	5560.0767	1.505	.019	3.659	.039
Dist_Bus_m	15836	169.9639	6.202	.019	60.039	.039
Dist_Railway_m	15836	847.3460	1.192	.019	1.816	.039
Dist_School_m	15836	218.7092	1.950	.019	5.867	.039
Walk_Index	15836	77.5980	-.915	.019	.271	.039
Valid N (listwise)	15836					

(Source: Developed from field data using SPSS)

The first operations to build the REP_benchmark starting from the REs-ZH-Dataset are the conversion of the variables, the calculation of the mean and the computation of the coefficient of variation for each variable. These operations do not necessarily require the assumption of a normal distribution. Thus, although the variables seem to not have normal distributions, no transformations with the goal to reach a normal distribution are done. Only the conversion of the variable “Age” is computed as the difference between the year 2007, which is the point of the cross-sectional analysis in this study, and the construction year of each RE. Details about the calculated statistical values for each of the 17 variables in the REs-ZH-Dataset are presented in Appendix 8.5.

5.3.4 Descriptive Analysis for REs-ZH-Dataset, N=15,836

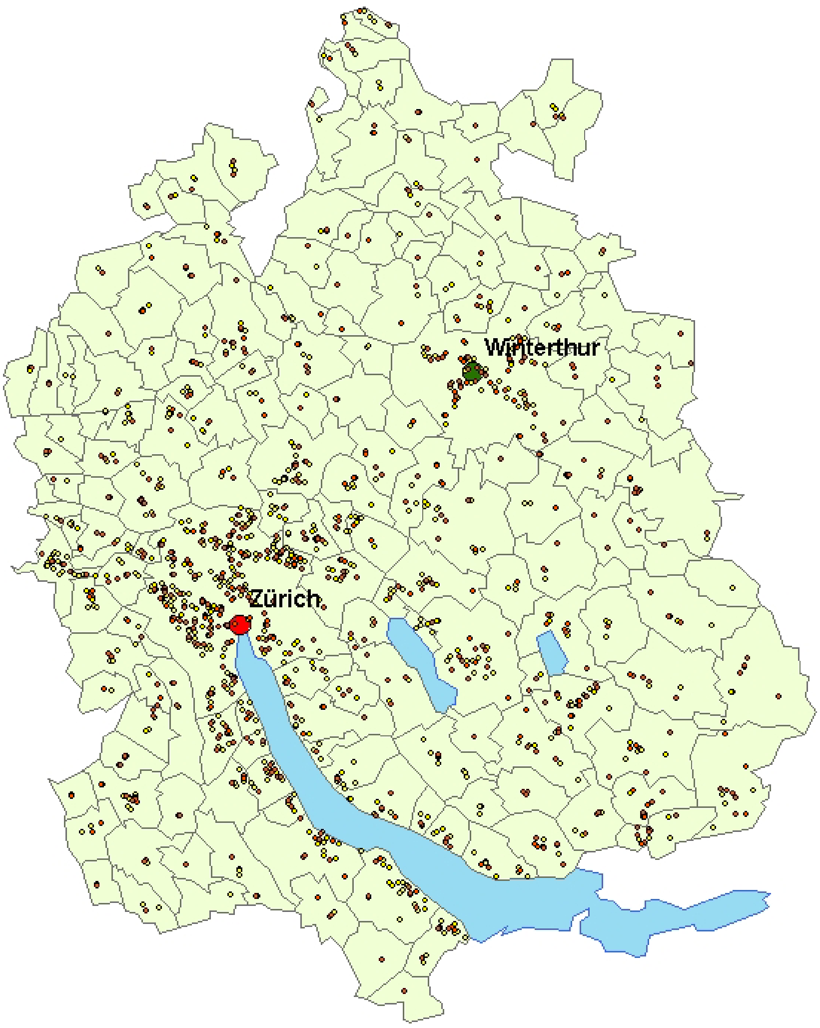
In this section, other interesting aspects of the data are highlighted, not with the goal of building all possible statistics over all variables included in the dataset but rather to better understand the REs-ZH-Dataset. Figure 5.4 shows a graphical distribution of the REs owned by the PFs included in the REs-ZH-Dataset. The distance to a centre is marked with orange dots for REs having the shortest distance to Zurich and with green dots for REs having the shortest distance to Winterthur. Zurich and Winterthur are the two largest centres in Canton Zurich (STA 2009). Paradeplatz (Appendix 8.7) is considered the centre point for Zurich and the railway station (Appendix 8.8) is considered the centre point for Winterthur. The geographical disposition of the REs seems concentrated around the two centres and in locations near to a lake.

Figure 5.4 Geographical Distribution of the PF owned REs in Canton Zurich

(Source: Developed from field data using GIS)

The next Figure 5.5 complements the Figure 5.4 with the geographical distribution of schools and nursery schools in order to show their density in Canton Zurich and their distances to the REs owned by the PFs.

Figure 5.5 Geographical Distribution of School and Nursery School in Canton Zurich



(Source: Developed from field data using GIS)

The next aspect of the data analysed refers to the variable "NrRooms" considered in Table 5.13 and that has been grouped using the variable "Usage" as factor in order to analyse the two independent groups SFD and CO separately. A result indicates that the SFD mean (4.25, N=110) is higher than the CO mean (3.16, N=15,726).

Table 5.13 Descriptive Statistic for Variables "NrRooms" with "Usage" as Factor

a) Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Usage	N	Percent	N	Percent	N	Percent
NrRooms	1.00=SFD	110	100.0%	0	.0%	110	100.0%
	2.00=CO	15726	100.0%	0	.0%	15726	100.0%

b) Descriptive Statistic

	Usage			Statistic	Std. Error
NrRooms	1.00=SFD	Mean		4.2545	.11726
		95% Confidence Interval for Mean	Lower Bound	4.0221	
			Upper Bound	4.4870	
		5% Trimmed Mean		4.2626	
		Median		5.0000	
		Variance		1.513	
		Std. Deviation		1.22988	
		Minimum		1.00	
		Maximum		8.00	
		Range		7.00	
		Interquartile Range		2.00	
		Skewness		-.139	.230
		Kurtosis		-.069	.457
	2.00=CO	Mean		3.1638	.00847
		95% Confidence Interval for Mean	Lower Bound	3.1472	
			Upper Bound	3.1804	
		5% Trimmed Mean		3.1709	
		Median		3.0000	
		Variance		1.127	
		Std. Deviation		1.06154	
		Minimum		1.00	
		Maximum		8.00	
		Range		7.00	
		Interquartile Range		1.00	
		Skewness		-.139	.020
		Kurtosis		-.075	.039

(Source: Developed from field data using SPSS)

The variable "Surface" considered in Table 5.14 has been grouped with the variable "NrRooms" as factor in order to analyse the independent groups for each number of rooms separately. The mean surface for REs with two rooms is 57.68 square metres with a standard deviation of 11.60. The mean surface for REs with three rooms is 74.08 with a standard deviation of 14.03. REs with three rooms have with a quote of 37.40 percent (N=5,923) the highest number of presences in the dataset, followed by 30.45 percent for REs with four rooms (N=4,822), by 16.01 percent for REs with two rooms (N=2,535), by 7.67 percent for REs with one room (N=1,214), by 7.52 percent for REs with five rooms (N=1,191) and the rest by 0.95 percent for REs with more than five rooms (151). The complete descriptive statistics are presented in Appendix 8.9.

Table 5.14 Descriptive Statistic for Variables "Surface" with "NrRooms" as Factor

a) Case Processing Summary

		Cases					
		Valid		Missing		Total	
(N=15836)	NrRooms	N (%)	Percent	N	Percent	N	Percent
Surface_m2	1.00	1214 (7.67)	100.0%	0	.0%	1214	100.0%
	2.00	2535 (16.01)	100.0%	0	.0%	2535	100.0%
	3.00	5923 (37.40)	100.0%	0	.0%	5923	100.0%
	4.00	4822 (30.45)	100.0%	0	.0%	4822	100.0%
	5.00	1191 (7.52)	100.0%	0	.0%	1191	100.0%
	6.00	133 (0.84)	100.0%	0	.0%	133	100.0%
	7.00	16 (0.10)	100.0%	0	.0%	16	100.0%
	8.00	2 (0.01)	100.0%	0	.0%	2	100.0%

b) Descriptive Statistic

	NrRooms			Statistic	Std. Error
Surface_m2	2.00	Mean		57.6852	.23051
		95% Confidence Interval for Mean	Lower Bound	57.2332	
			Upper Bound	58.1372	
		Median		56.0000	
		Std. Deviation		11.6059	
		Skewness		1.110	.049
		Kurtosis		7.015	.097
	3.00	Mean		74.0848	.18239
		95% Confidence Interval for Mean	Lower Bound	73.7272	
			Upper Bound	74.4423	
		Median		74.0000	
		Std. Deviation		14.0366	
		Skewness		1.120	.032
		Kurtosis		5.482	.064

The highly significant ($p < 0.01$) bivariate correlation of 0.813 between the variable "Surface" and "NrRooms" (Table 5.15) is confirmed by the box-plot in the Figure 5.6 that indicates an increment of the surface with an increment of the number of rooms.

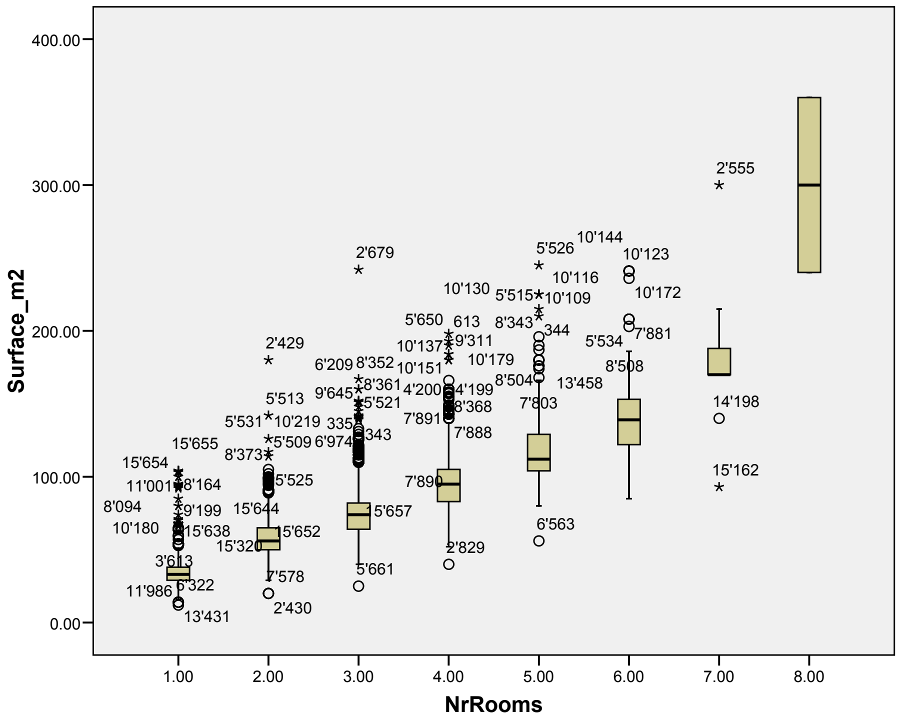
Table 5.15 Pearson's Correlation between Variables "Surface" and "NrRooms"

		NrRooms	Surface_m2
NrRooms	Pearson's Correlation	1	.813(*)
	Sig. (2-tailed)		.000
Surface_m2	Pearson's Correlation	.813(*)	1
	Sig. (2-tailed)	.000	

* Correlation is significant at the 0.01 level (2-tailed).
a Listwise N=15836

(Source: Developed from field data using SPSS)

Figure 5.6 Boxplot Variables "Surface" with "NrRooms" as Factor



(Source: Developed from field data using SPSS)

The variable "DistanceToCentre" is depicted in Table 5.16. For 50 percent of the REs included in the REs-ZH-Dataset, the distance to a centre is less than 7.0 kilometres, for 75 percent of them the distance is less than 11.1 kilometres and for 90 percent of them the distance is less than 17.7 kilometres.

Table 5.16 Frequency Summary Variable "DistanceToCentre_m"

		Frequency	Percent	Valid Percent	Cum. Percent
	…	…			
Valid	2067.71	7	0.04	0.04	10.00
	…	…			
	6682.74	20.0	0.12	0.12	50.07
	…	…			
	11021.37	18	0.11	0.11	75.08
	…	…			
	17678.76	8	0.05	0.05	90.04
	…	…			
	29848.58	8	0.05	0.05	100.00
	Total	15836	100	100	

(Source: Developed from field data using SPSS)

The variable "Age" depicted in Table 5.17 shows that for 51.48 percent of the REs in the REs-ZH-Dataset the age is less than 35 years old, for 76.09 percent of them the age is less than 47 years old and for 90.3 percent of them the age is less than 56 years old.

Table 5.17 Frequency Summary Variable "Age"

		Frequency	Percent	Valid Percent	Cum. Percent
	…	…			
Valid	11	140	0.88	0.88	10.31
	…	…			
	35	541	3.41	3.41	51.48
	…	…			
	47	356	2.24	2.24	76.09
	…	…			
	56	131	0.82	0.82	90.30
	…	…			
	88	365	2.30	2.30	100.00
	Total	15836	100	100	

(Source: Developed from field data using SPSS)

In this section the raw data of the REs-ZH-Dataset has been screened in order to increase the data quality. The explored data contained twelve missing values that have been imputed and 1,290 REs samples with evident errors and outliers that have been eliminated from the analysis resulting in a reduction of the REs-ZH-Dataset samples from 76 PFs to 74 PFs. No transformations have been applied on the 17 variables. The exception is the variable “Age” that has been converted from an absolute value to a relative value. At the end, additional descriptive characteristics of the REs-ZH-Dataset have been presented. Now that this data screening and analysis has been performed, the next section presents the computation of the REP_benchmark. This is an important element of this study.

5.4 Computation of the PF’s REP_benchmark (N=74)

This section begins the most important part of the analysis: *the computation of the REP_benchmark*. The REs-ZH-Dataset is prepared using a crisscross table to extract from each of the 74 PFs the statistical values (mean and standard deviation) of their variables that are needed in order to normalise the variables and compute the PF’s REP_benchmark (Appendix 8.10). After the dataset is prepared, the analysis will follow the procedure defined in the methodology Chapter 4, which includes following steps: a) The coefficients of variation (*cv*) for the variables and for the 74 PFs are calculated using the Formula 4.2 and used in Formula 4.3 for the determination of the *cv* benchmarks for each variable; b) the categorical variable “Usage” is converted into a metrical variable using the Formula 4.4 and the *cv* benchmark for this variable is calculated with the

Formula 4.5 and c) the linear normalisation is applied using the Formula 4.7 between the mean of each PF's variable (μ_{ij}) and the mean of all PFs' means for the same variable (μ_{iB}), where the minimum value gets transformed to zero and the μ_{iB} is set to one.

When the computation of the PFs REP_benchmark (N=74) is undertaken according to the procedure outlined in Chapter Four and summarised in the previous paragraph, the results generated are depicted in Table 5.18. This table contains normalised variables ($NormV_{iB}$), coefficients of variation (cv_{iB}) and means of the variables for each PF (μ_{iB}).

Table 5.18 Summary of the PFs REP_benchmark, N=74

	REP_benchmark values between all PFs					
	$NormV_{iB}$		Mean		Coefficient of Variation	
Predictors	N	$NormV_{iB}$	N	$Mean_{iB}$	N	cv_{iB}
Usage	74	1.00	74	2.3214	74	4.0613
Age	74	1.00	74	43.1383	74	0.3008
NrRooms	74	1.00	74	3.3712	74	0.2967
Surface_m2	74	1.00	74	86.7942	74	0.2754
Volume_m3	74	1.00	74	5329.9376	74	0.3345
GeoLocationX_m	74	1.00	74	687907.4814	74	0.0036
GeoLocationY_m	74	1.00	74	248762.3524	74	0.0091
DistanceToCentre_m	74	1.00	74	9956.1022	74	0.2366
TaxationLevel_Percent	74	1.00	74	122.0860	74	0.0367
Lake_ha	74	1.00	74	783.0031	74	0.7754
View_ha	74	1.00	74	15806.0937	74	0.3060
Slope_Percent	74	1.00	74	3.5815	74	0.4168
SunJuly_KJ_m2	74	1.00	74	5584.1485	74	0.0112
Dist_Bus_m	74	1.00	74	174.3269	74	0.3007
Dist_Railway_m	74	1.00	74	869.3862	74	0.2585
Dist_School_m	74	1.00	74	244.2774	74	0.3440
Walk_Index	74	1.00	74	76.4104	74	0.1103

(Source: Developed for this research)

5.5 "Distance" between a PF's REP and the PFs' REP_benchmark

Using the computed PFs REP_benchmark, it is now possible to calculate the multivariate "distance" between the PFs REP_benchmark (REP_B) and the 74 PF's REP (REP_S) contained in the REs-ZH-Dataset. The computation is undertaken using the Formula 4.9 and the result is saved into a new sub-dataset called *PFs-ZH-Dataset* that contains, for each of the 74 PFs, a dependent variable (the Euclidean distance) and independent variables (the percentage values for the variable "Usage" and the mean values for the other variables as summarised in Table 5.19).

Table 5.19 Summary of the *PFs-ZH-Dataset* with Euclidean Distance, N=74

	Dependent (Y)	Independent (Xi)					
PF_ CodeNr	EuclideanDist	Usage percentage	Age Mean	NrRooms Mean	Surface_m2 Mean	Volume_m3 Mean	… …
1.00	6.23	0.00	50.00	3.20	67.00	2 830.00	
2.00	1.44	0.00	32.56	3.55	76.41	4 389.59	
3.00	2.11	0.00	88.00	3.67	106.67	840.00	
4.00	5.64	9.09	38.42	3.75	113.33	2 710.75	
5.00	2.24	0.00	32.00	4.00	87.00	1 164.00	
6.00	2.18	0.00	88.00	3.67	86.44	3 950.00	
7.00	1.63	0.00	36.05	3.61	85.74	3'851.50	
…	…	…	…	…	…	…	
69.00	0.91	3.52	35.66	3.08	87.66	6 982.82	
70.00	1.47	0.00	37.00	2.19	53.13	1 964.50	
71.00	2.00	0.00	39.87	2.97	63.50	9 343.67	
72.00	3.22	9.57	42.97	2.83	76.81	5 846.62	
73.00	2.33	0.00	56.00	3.67	75.33	5 575.00	
74.00	2.08	0.00	34.00	3.50	76.00	7 895.00	

(Source: Developed for this research)

The Euclidean distance is the dependent variable that has been calculated and included into the PFs-ZH-Dataset and the resulting "distance" expressed with this variable "EuclideanDist" ranges from a minimum of 0.774 to a maximum of 31.369 with a mean value of 2.923 indicating the presence of possible outliers. Outliers could be a problem when using the multivariate regression method (Chapter 4) to derive the decision model (REP-EDM) of this study, therefore a deeper data analysis of this new data set (PFs-ZH-Dataset) is undertaken in the next section.

5.6 *PFs-ZH-Dataset* - Data Screening, Transformation and Selection

The PFs-ZH-Dataset is screened before proceeding with the development of the REP-EDM, which is based on a multivariate regression analysis. The techniques used involve an examination of the individual variables and the relationships among them. The first is done with the calculation of values such as the mean, range and standard deviation and the second with the calculation of correlations and using graphical diagrams such as the scatterplot and box-plot. The graphical techniques are meant as a complementary means of portraying the data and its relationship.

5.6.1 Outliers Analysis

Errors and missing values have already been corrected and imputed in the REs-ZH-Dataset and, while the PFs-ZH-Dataset - without the added dependent variable "EuclideanDist" - is a sub dataset of it. Thus, no errors or missing values are present. The Euclidean distance has been calculated for each sample then added as dependent variable to the PFs-ZH-Dataset, therefore an analysis for outliers is necessary before the use of a multivariate regression analysis. Hair et al. (2006, p. 76) affirm that outliers '... should be retained unless demonstrable proof indicated that they are truly aberrant and not representative of any observations in the population.' According to Statsoft (2009, p. 6) 'there is no widely accepted method to remove outliers automatically and some researchers use quantitative methods to exclude outliers, they exclude observations that are outside the range of ±2 standard deviation.'

After an explorative analysis on the PFs-ZH-Dataset two outliers (PF number 38 and 55) are eliminated because their values on various variables are outside the range of ±2 standard deviation. The variable "EuclideanDist" has a mean value of 2.92 (N=74) and a standard deviation of 3.68 (N=74). The calculated range that defines the outliers (mean ± 2*standard deviation = 2.92 ± 7.37) starts by −4.45 until +10.29 and the PF Nr. 55 has a "distance" of 31.37, which is outside the admitted range, therefore it is omitted from the PFs-ZH-Dataset. In fact, the value of 31.37 corresponds to a factor of about 300 percent of the upper border of the outliers range (Appendix 8.11). The PF Nr. 38 is omitted for the same reason. In the end, the PFs-ZH-Dataset is composed by 72 PFs with one dependent variable "EuclideanDist" and 17 independent variables. Another important issue for a multivariate regression analysis is the data transformation. This is discussed in the next section.

5.6.2 Data Transformations

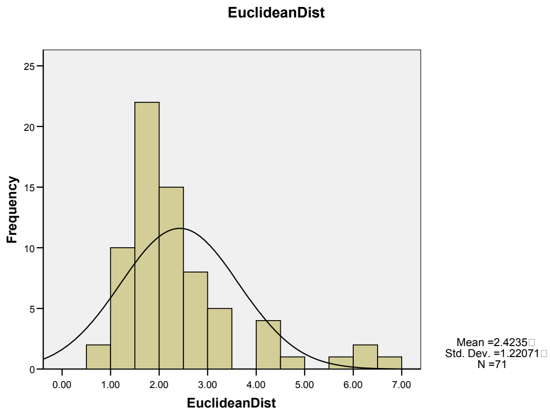
According to Coakes and Steed (2007), transformation of some variables may be used to attain normality. If variables display non-normal distribution, they can be transformed before further analysis. The decision to transform them depends on various factors such as the severity of the departure from normality, the statistical method used (parametric or non-parametric) and whether a multivariate analysis violates its assumption, such as showing a non-normal distribution of its residuals.

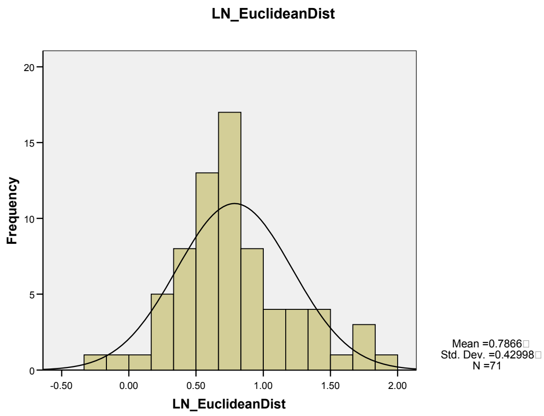
The results of the residual analysis can be examined only at the end of the regression modelling process. Although not all variables present a normal distribution (Appendix 8.12), they are not transformed per default but possible transformations of the variables are considered in the iterative process during the development of the regression model. In order to facilitate the selection process of the best regression equation, only the dependent variable "EuclideanDist" which originally presents outliers and a non-normal distribution is prepared with a transformation that aims to normalise it. The natural logarithm is used to transform this variable into "LN_EuclideanDist" which presents a better normal distribution as confirmed by the test of normality, histogram, boxplot and Q-Q plot (Table 5.20, Figure 5.7).

Coakes and Steed (2007) describe different ways to explore the normality assumption graphically with diagrams such as histograms, steam-and-leaf plots, boxplots and normal probability plots, and statistically with such tests as Kolmogorov-Smirnov statistic, Lilliefors significance level and the Shapiro-Wilk statistic. Different tests have been undertaken in order to test this assumption. According to the following histogram, the boxplots and the Q-Q plot the distribution of the transformed variable "LN_EuclideanDist" appears to be approximately normal.

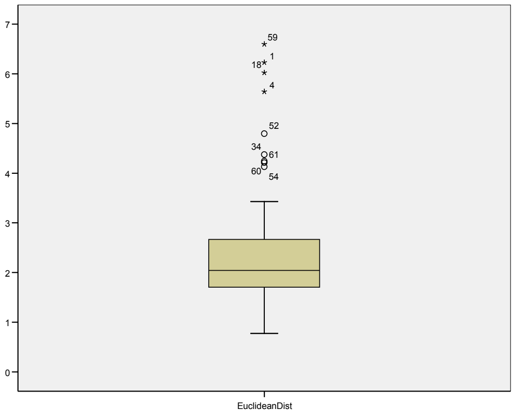
Figure 5.7 Graphical Analysis - Variables “EuclideanDist”, LN_EuclideanDist”

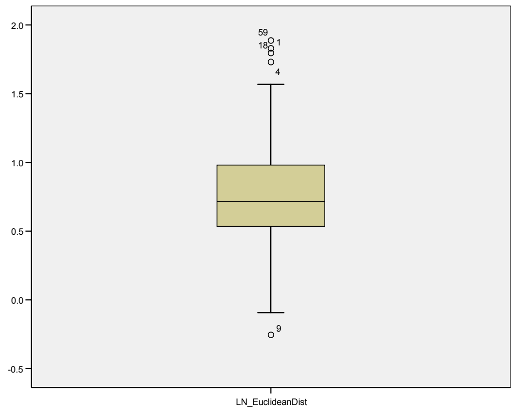
a) Histograms



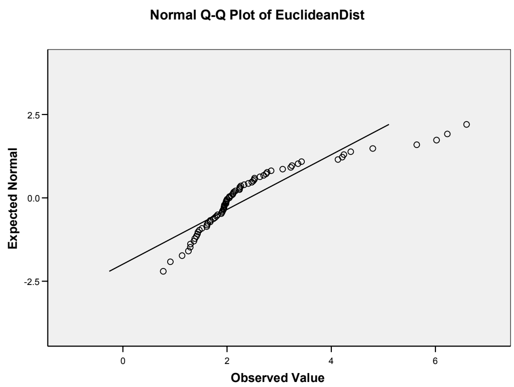


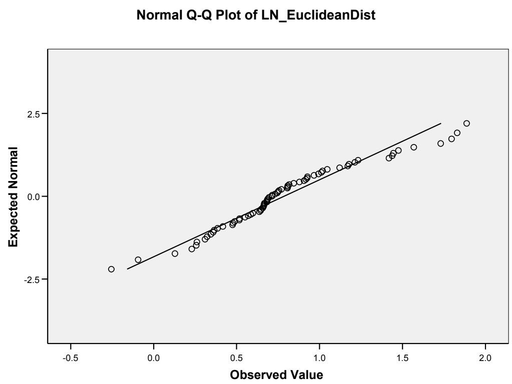
b) Boxplots





c) Normal Q-Q Plot





(Source: Developed from field data using SPSS)

In Table 5.20a, the test SW seems to confirm the graphical conclusion, which the distribution of "LN_EuclideanDist" appears to be normal. In Table 5.20b, the Skewness (0.559) and the Kurtosis (0.629) do not demonstrate evidence of a violation of the normality assumption.

Table 5.20 Descriptive Statistic - Variables "EuclideanDist", "LN_EuclideanDist"

a) Tests of Normality

(N=71)	Kolmogorov-Smirnov(a) (KS)			Shapiro-Wilk (SW)		
	Statistic	df	Sig.	Statistic	df	Sig.
EuclideanDist	.200	71	.000	.808	71	.000
LN_EuclideanDist	.120	71	.013	.959	71	.020

a Lilliefors Significance Correction

b) Descriptive Statistic

			Statistic	Std. Error
EuclideanDist	Mean		2.4235	.14487
	95% Confidence Interval for Mean	Lower Bound	2.1345	
		Upper Bound	2.7124	
	Std. Deviation		1.22071	
	Skewness (SK)		1.796	.285
	Kurtosis (KU)		3.164	.563
LN_EuclideanDist	Mean		.7866	.05103
	95% Confidence Interval for Mean	Lower Bound	.6848	
		Upper Bound	.8884	
	Std. Deviation		.42998	
	Skewness (SK)		.559	.285
	Kurtosis (KU)		.629	.563

In brief, although the Lilliefors statistic suggests that there is still a slight problem, all of the other diagnostics of the data are satisfactory. Thus, from the above statistics and graphs it seems that the natural logarithmic transformation for the variable "EuclideanDist" is appropriate because the distribution of the variable "LN_EuclideanDist" is now normal.

5.6.3 Random Selection of a PF for Discussion and Practical Analysis

According to the methodology procedure (Chapter 4.10) one of the 72 PFs contained in the PFs-ZH-Dataset has to be randomly selected and extracted from the dataset. The aim is to develop a regression model independently of this selected PF. The randomly chosen PF has the number 6, thus it is extracted from the PFs-ZH-Dataset, which contains now 71 PFs. This PF will be later introduced to the model for analysis and discussions of results in Chapter Six (out of sample analysis).

5.7 Summary of the PFs REP_benchmark (N=71) used for REP-EDM

In the previous section the PFs-ZH-Dataset has been screened with the aim to increase the data quality and to improve the result of the multivariate regression. The explored data contained two outliers that have been eliminated from the analysis resulting in a reduction of the PFs-ZH-Dataset samples from 74 PFs to 72 PFs. The dependent variable "EuclideanDist" has been transformed with a natural logarithm in order to prepare data for the iterative selection process of the best regression equation. At the end, one PF has been extracted to be used in the discussion chapter and the final resulting PFs REP_benchmark (Table 5.21) used to derive the decision-model (REP-EDM) contains 71 PFs and this is inspected more in depth with a descriptive analysis in the next section.

Table 5.21 Summary of the PFs REP_benchmark used for REP-EDM, N=71

	REP_benchmark values between all PFs					
	$NormV_{iB}$		Mean		Coefficient of Variation	
Predictors	N	$NormV_{iB}$	N	$Mean_{iB}$	N	cv_{iB}
Usage	71	1.00	71	.9925	71	2.6929
Age	71	1.00	71	42.2882	71	0.3002
NrRooms	71	1.00	71	3.3614	71	0.2931
Surface_m2	71	1.00	71	86.3683	71	0.2726
Volume_m3	71	1.00	71	5420.7801	71	0.3378
GeoLocationX_m	71	1.00	71	687531.31	71	0.0037
GeoLocationY_m	71	1.00	71	248825.17	71	0.0093
DistanceToCentre_m	71	1.00	71	9929.4531	71	0.2448
TaxationLevel_Percent	71	1.00	71	121.6335	71	0.0381
Lake_ha	71	1.00	71	792.1803	71	0.7709
View_ha	71	1.00	71	16068.826	71	0.3087
Slope_Percent	71	1.00	71	3.5881	71	0.4134
SunJuly_KJ_m2	71	1.00	71	5580.8610	71	0.0115
Dist_Bus_m	71	1.00	71	170.6932	71	0.3048
Dist_Railway_m	71	1.00	71	884.1523	71	0.2578
Dist_School_m	71	1.00	71	243.9126	71	0.3413
Walk_Index	71	1.00	71	76.6955	71	0.1131

(Source: Developed for this research)

5.8 *PFs-ZH-Dataset (N=71)* - Descriptive Analysis for REP-EDM

The first aspect analysed is depicted in the Table 5.21 which shows the PFs REP_benchmark with the mean of the variables in the PFs-ZH-Dataset. The mean values of the PFs REP_benchmark are not the same as those calculated for the variables in the REs-ZH-Dataset. For example, the mean of the variables “Age” diverges by 19.2 percent, the mean of the variable “NrRooms” diverges by 5.9 percent and the variable “Surface” diverges by 9.9 percent. These differences in the mean values between the REs-ZH-Dataset and the PFs-ZH-Dataset are the result of the calculation of the PF’s REP_benchmark that neutralises the size (number of REs) of the PF. Thus the PF’s REP_benchmark is independent of the PF size and its investment strategy. For example, a

PF that owns a REP including 1,000 REs and that applies a luxury strategy of RE investment could, with its potential for a high frequency of extreme values compared to another PF that has only 10 REs in its REP and that invests in common REs, over-influence the REP_benchmark. Computing the mean of each PF, its REP size is neutralised. The detailed descriptive and explorative analyses for each variables of the PFs-ZH-Dataset are available in Appendix 8.13.

The second aspect analysed includes the tests of normality on the variables (Table 5.22).

Table 5.22 Tests of Normality for Variables in the *PFs-ZH-Dataset*, N=71

a) Kolmogorov-Smirnov and Shapiro-Wilk

	Kolmogorov-Smirnov(a) (KS)			Shapiro-Wilk (SW)		
	Statistic	df	Sig.	Statistic	df	Sig.
EuclideanDist	.200	71	.000	.808	71	.000
LN_EuclideanDist	.120	71	.013	.959	71	.020
Usage	.391	71	.000	.436	71	.000
Age	.140	71	.002	.943	71	.003
NrRooms	.157	71	.000	.884	71	.000
Surface_m2	.135	71	.003	.920	71	.000
Volume_m3	.144	71	.001	.801	71	.000
GeoLocationX_m	.087	71	.200(*)	.977	71	.216
GeoLocationY_m	.096	71	.171	.963	71	.035
DistanceToCentre_m	.153	71	.000	.939	71	.002
TaxationLevel_Percent	.155	71	.000	.877	71	.000
Lake_ha	.261	71	.000	.684	71	.000
View_ha	.093	71	.200(*)	.964	71	.042
Slope_Percent	.180	71	.000	.818	71	.000
SunJuli_KJ_m2	.169	71	.000	.865	71	.000
Dist_Bus_m	.197	71	.000	.625	71	.000
Dist_Railway_m	.203	71	.000	.769	71	.000
Dist_School_m	.175	71	.000	.867	71	.000
Walk_Index	.177	71	.000	.856	71	.000

* This is a lower bound of the true significance.
a Lilliefors Significance Correction

b) Skewness and Kurtosis

	N	Skewness		Kurtosis	
Valid N = 71 (listwise)	Statistic	Statistic	Std. Error	Statistic	Std. Error
EuclideanDist	71	1.796	.285	3.164	.563
LN_EuclideanDist	71	.559	.285	.629	.563
Usage	71	3.571	.285	13.989	.563
Age	71	.539	.285	.288	.563
NrRooms	71	1.247	.285	4.018	.563
Surface_m2	71	1.031	.285	1.460	.563
Volume_m3	71	2.294	.285	7.796	.563
GeoLocationX_m	71	.191	.285	-.358	.563
GeoLocationY_m	71	.650	.285	1.267	.563
DistanceToCentre_m	71	.533	.285	-.595	.563
TaxationLevel_Percent	71	-1.247	.285	1.137	.563
Lake_ha	71	2.244	.285	5.240	.563
View_ha	71	.515	.285	1.701	.563
Slope_Percent	71	1.852	.285	4.022	.563
SunJuli_KJ_m2	71	1.647	.285	4.726	.563
Dist_Bus_m	71	4.492	.285	28.600	.563
Dist_Railway_m	71	2.819	.285	12.789	.563
Dist_School_m	71	1.848	.285	5.896	.563
Walk_Index	71	-1.714	.285	4.088	.563

(Source: Developed from field data using SPSS)

The tests of normality on the variables indicates the presence of outliers and various non-normal distributions. This is demonstrated on the Table 5.22 by the tests KS and SW. The significance of KS is smaller than 0.05 for most variables and for various variables SW is too far below 1.0. Additionally, Skewness and Kurtosis demonstrate evidence of a violation of the normality assumption by various variables.

The next aspect examined considers the number of REs that are owned by each single PF and the corresponding question is: How many REs does each PF own? The answer to this question can be found in the Table 5.23. The first seven PFs own more than 76 percent of the REs owned by PFs in Canton Zurich. The smallest PF REP contains only one RE. The complete table is available in Appendix 8.14.

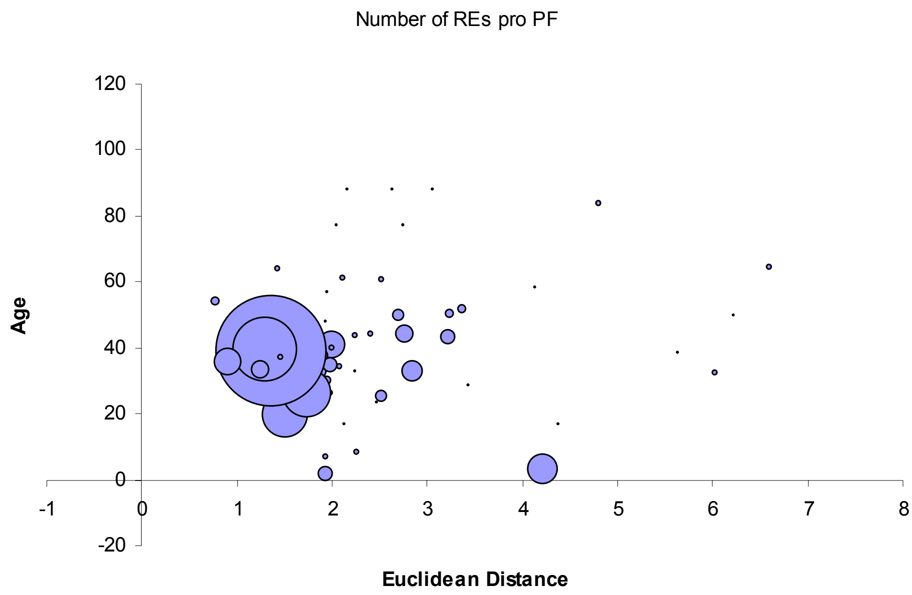
Table 5.23 Number of REs per PFs in Canton Zurich

PF_CodeNr	Number REs	Percent
66	5022	31.79%
12	1836	11.62%
67	1674	10.60%
39	1275	8.07%
58	989	6.26%
26	839	5.31%
64	440	2.79%
69	353	2.23%
48	349	2.21%
23	258	1.63%
2	213	1.35%
51	208	1.32%
…	…	…
Total	15796	100.00%
Mean REs pro PF (N=71)	222.48	-

(Source: Developed from field data using SPSS)

Another aspect of the data analysed refers to the relationship between the variables “EuclideanDist”, “Age” and the number of REs pro PF, as depicted in Figure 5.8. The distribution seems to be concentrated around the mean values of the single variables: For the variable “Age” around its mean of 42.28 years old, and for the variable “EuclideanDist” around its mean of 2.42. PFs that own a high number of REs seem to be concentrated on these averages.

Figure 5.8 Relationship "EuclideanDist", "Age" and Number of REs pro PF



(Source: Developed from field data using SPSS)

The last aspect analysed is the relationship depicted in Table 5.24 and describes the bivariate correlations between the dependent variable "EuclideanDist" and the independent variables. The correlation of 0.556 with the variable "Lake" and of 0.504 with the variable "Usage" are very highly significant ($p < 0.001$). The correlation of 0.363 with the variable "DistanceToCentre", the correlation of 0.327 with the variable "Slope" and of 0.304 with the variable "Surface" are highly significant ($p < 0.01$). The next two variables "GeoLocation" and "NrRooms" indicate a significant ($p < 0.05$) correlation and for the rest of the variables the correlation is not significant.

Table 5.24 Bivariate Correlation Pearson's Correlation (X Variables to Y)

(N 71)	**EuclideanDist (Y)**	Sig. (2-tailed, significant 0.05)	Ranking Influence (best =1)
Lake_ha	0.555707677	0.000000489	1
Usage	0.504583484	0.000007227	2
DistanceToCentre_m	0.363367894	0.001841621	3
Slope_Percent	0.327263167	0.005339944	4
Surface_m2	0.304749821	0.009764184	5
GeoLocationY_m	-0.286682461	0.015356318	6
NrRooms	0.233080054	0.050448629	7
Dist_Bus_m	0.211613778	0.076467356	8
Dist_Railway_m	0.181855147	0.129058027	9
Walk_Index	-0.178070445	0.137353795	10
SunJuly_KJ_m2	-0.108720826	0.366782603	11
Age	0.088068226	0.465191541	12
GeoLocationX_m	0.021841945	0.856525927	13
Dist_School_m	0.021359316	0.859663648	14
TaxationLevel_Percent	0.013773909	0.909232786	15
View_ha	-0.003915331	0.974148528	16
Volume_m3	0.000405717	0.997320749	17

(Source: Developed from field data using SPSS)

5.9 Development of the REP Empirical Decision Model (REP-EDM)

The variables in the PFs-ZH-Dataset (N=71) have been converted and transformed with the goals of preparing the data facilitating the selection process of the best regression equation, of creating a unit-less “distance” to the PFs REP_benchmark and because ‘…the regression analyses the relationship between changes among variables, rather than the absolute levels of the variables’ (Studenmund 2006, p. 402). The goals for using the multivariate regression on the prepared PFs-ZH-Dataset are to analyse and quantify the relationship between the dependent variable “EuclideanDist” (liquidity risk) and various independent variables such as location, taxes, distance to the centre and to develop a model that provides the best prediction of Euclidean distances in the area of the canton of Zurich based on the PFs-ZH-Dataset.

5.9.1 Selection of the "Best" Regression Equation

After the raw data has been analysed, the selection of the "best" prediction multiple linear regression model can begin. According to Studenmund (2006, p.390) '…we never know what the true model is', in addition Gujarati (2003, p. 74) affirms that '…to some extent there is some trial and error involved in choosing the "right" model for empirical analysis.' Despite the difficulty of not having only a single procedure to follow, there are numerous different strategies for selecting the "best" model when the primary goal of analysis is prediction. Kachigan (1991) indicates following two procedures to reduce and optimise the number of variables into a regression: 1) Stepwise procedures with forward addition or backward elimination and 2) All regressions test that tests every possible regression equations through a computer simulation. Kleinbaum et al. (1998) include these procedures into a five-step process that is used to select the model in this study.

<u>1. Step: Specify the maximum model to be considered</u>

The model estimation starts with the maximum model that includes the dependent variable "EuclideanDist" and the 17 independent variables according to Appendix 8.1.

<u>2. Step: Specify a criterion for selecting a model</u>

The criterion is a trade between the higher "Adjusted R^2" and an acceptable homoscedasticity or "estimated error variance" under the condition that the ANOVA F-test testing the overall significance of the model is significant. A significant F-test allows the rejection of the null hypothesis (H_0) that all coefficients are equal to zero. In the chosen model all coefficients must be significant.

3. Step: Specify a strategy for selecting variables

The level of correlation between regressors and regressand (Table 5.32) together with a significant t-test of the regressor's coefficient determine which variables are chosen.

4. and 5. Steps: Conduct a specified analysis, evaluate the reliability of the model chosen

In order to select the predictors for the "best" multiple regression estimations (MRE) an iterative process is applied and the following estimations are executed.

a) MRE with All Predictors Entered Simultaneously (ED)

The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.15.

Table 5.25 a) MRE Regression Equation Coefficients with Significance

Model	Unstandardised Coefficients		t	Sig.
	B	Std. Error		
(Constant)	4.880	10.463	.466	.643
Surface_m2	.012	.008	1.546	.128
Lake_ha	.001	.000	4.910	.000
Usage	.203	.037	5.452	.000
DistanceToCentre_m	3.15E-005	.000	1.629	.109
Dist_Railway_m	.000	.000	1.351	.182
Age	-.001	.006	-.229	.819
NrRooms	-.060	.212	-.284	.777
Volume_m3	1.82E-006	.000	.076	.940
GeoLocationX_m	-1.02E-005	.000	-.718	.476
GeoLocationY_m	4.91E-006	.000	.343	.733
TaxationLevel_Percent	.011	.009	1.277	.207
View_ha	4.11E-006	.000	.259	.796
Slope_Percent	.018	.045	.390	.698
SunJuli_KJ_m2	7.40E-005	.001	.084	.934
Dist_Bus_m	-7.57E-005	.001	-.075	.940
Dist_School_m	5.82E-005	.001	.069	.945
Walk_Index	-.008	.007	-1.082	.284

(Source: Developed from field data using SPSS)

The Table 5.25 indicates the regression equation coefficients with their significance for the 17 independent variables entered simultaneously into the model that uses the "EuclideanDist (ED)" as dependent variable. The R^2 and adjusted R^2 are respectively 0.856 and 0.646. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are "Lake_ha" and "Usage." Almost all coefficients' variables are not significant ($p > 0.05$) therefore this model is not optimal. The next step is to estimate a model with the same variable but using the stepwise entry procedure.

b) MRE with All Predictors Entered Stepwise (ED)

The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.16.

Table 5.26 b) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		T	Sig.
		B	Std. Error		
5	(Constant)	.149	.361	.414	.680
	Lake_ha	.001	.000	8.511	.000
	Usage	.212	.031	6.820	.000
	DistanceToCentre_m	3.83E-005	.000	2.520	.014
	Surface_m2	.011	.004	2.726	.008
	Dist_Railway_m	.000	.000	2.012	.048

(Source: Developed from field data using SPSS)

The Table 5.26 indicates the regression equation coefficients with their significance for the independent variables entered stepwise into the model that uses the "EuclideanDist (ED)" as dependent variable. The R^2 and adjusted R^2 are respectively 0.841 and 0.685.

The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are “Lake_ha,” “Usage,” “DistanceToCentre,” “Surface,” and “Dist_Railway.” The constant coefficient is not significant ($p > 0.05$) and this MRE removed 12 variables and included 5 variables. The resulting model presents a higher adjusted R^2 (0.685) than the model (a), but the constant coefficient is not significant. The next step is to exclude the constant from the MRE.

c) MRE with All Predictors Entered Stepwise (ED, no Constant)

The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.17.

Table 5.27 c) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		t	Sig.
		B	Std. Error		
5	Surface_m2	.012	.002	5.552	.000
	Lake_ha	.001	.000	8.864	.000
	Usage	.211	.031	6.851	.000
	DistanceToCentre_m	3.87E-005	.000	2.568	.013
	Dist_Railway_m	.000	.000	2.190	.032

(Source: Developed from field data using SPSS)

The Table 5.27 indicates the regression equation coefficients with their significance for the independent variables entered stepwise into the model that uses the “EuclideanDist (ED)” as dependent variable and excludes the constant variable. The R^2 and adjusted R^2 are respectively 0.970 and 0.937. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are “Lake_ha,” “Usage,” "DistanceToCentre," "Surface," and "Dist_Railway." The no multicollinearity assumption seems to be violated with a VIF > 5 and the resulting model presents a higher adjusted R^2 (0.937) than the model (b). This MRE removed 12 variables and included 5 variables. Three variables are very highly significant ($p < 0.001$) and two are significant ($p < 0.05$). The problem with this model is the multicollinearity assumption seems to have been violated. The next step is to include in the MRE only the five significant variables and using the transformed "LN_EuclideanDist" dependent variable.

d) MRE with Five Predictors Entered Simultaneously (LN_ED, no Constant)

The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.18.

Table 5.28 d) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		t	Sig.
		B	Std. Error		
1	Surface_m2	.004	.001	4.297	.000
	Lake_ha	.000	.000	6.065	.000
	Usage	.057	.014	4.155	.000
	DistanceToCentre_m	1.45E-005	.000	2.166	.034
	Dist_Railway_m	9.15E-005	.000	1.373	.174

(Source: Developed from field data using SPSS)

The Table 5.28 indicates the regression equation coefficients with their significance for the five independent variables entered simultaneously into the model that uses the "LN_EuclideanDist (LN_ED)" as dependent variable and excludes the constant variable. The R^2 and adjusted R^2 are respectively 0.946 and 0.887. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are

"Surface," "Lake_ha," "Usage," and "DistanceToCentre." The resulting model presents a lower adjusted R^2 (0.887) than the model (c) and not all coefficients are significant. In fact, the coefficient of the variable "Dist_Railway" is not significant ($p > 0.05$). The next step is to examine a MRE with the four selected variables.

e) MRE with Four Predictors Entered Simultaneously (ED, no Constant)

The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.19.

Table 5.29 e) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	Surface_m2	.015	.002	.489	7.522	.000	.222	4.503
	Lake_ha	.001	.000	.306	8.478	.000	.721	1.388
	Usage	.213	.032	.223	6.731	.000	.856	1.168
	DistanceToCentre_m	4.62E-005	.000	.196	3.062	.003	.230	4.351

(Source: Developed from field data using SPSS)

The Table 5.29 indicates the regression equation coefficients with their significance for the four independent variables entered simultaneously into the model that uses the "EuclideanDist (ED)" as dependent variable and excludes the constant variable. The R^2 and adjusted R^2 are respectively 0.968 and 0.933. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are "Surface," "Lake_ha," "Usage," and "DistanceToCentre." This model explains 93.3 percent (adjusted R^2) of the variance and does not present evidence of particular problems.

An examination of the t-values indicates that the variables “Surface“, “Lake_ha“ and “Usage“ are very highly significant ($p < 0.001$) and the variable “DistanceToCentre” is highly significant ($0.01 > p < 0.001$) and therefore all predictors significantly contribute to the prediction of the Euclidean distance. All the predictors have a positive coefficient influencing the Euclidean distance positively. Kachigan (1991, p. 184) affirms that “…the raw score form of the regression equation is fine for predicting actual values of the criterion variable *y*, but the beta coefficients from the standardised score form of the equation are needed to interpret the relative importance of the various predictor variables.” According to the standardised coefficients the major contribution to the model is given by the variable “Surface” with a beta of 0.489, the next is the variable “Lake” with a beta of 0.306, the next is the variable “Usage” with a beta of 0.223 and the last is the variable “DistanceToCentre” with a beta of 0.196. The next step is to use the curve fitting method with the aim to improve the MRE.

f) MRE with Curve Fitting of Variable (ED, no Constant)

Using the curve fitting of a variable instead of the variable itself could improve the MRE. In this case, various combinations have been tested and the combination with a cubic curve fitting for the variables “Surface” (Appendix 8.21) and “Lake” (Appendix 8.22) that reaches an adjusted R^2 of 0.914 is described. The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.20.

Table 5.30 f) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		t	Sig.
		B	Std. Error		
1	CF_Surface_m2	.500	.084	5.943	.000
	Usage	.213	.036	5.954	.000
	CF_Lake_ha	.329	.063	5.192	.000
	DistanceToCentre_m	5.90E-005	.000	3.586	.001

(Source: Developed from field data using SPSS)

The Table 5.30 indicates the regression equation coefficients with their significance for the four independent variables entered simultaneously into the model that uses the "EuclideanDist (ED)" as dependent variable and excludes the constant variable. The R^2 and adjusted R^2 are respectively 0.958 and 0.914. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are "CF_Surface," "CF_Lake_ha," "Usage," and "DistanceToCentre." This model does not present evidence of particular problems. The resulting model presents a lower adjusted R^2 (0.914) than the model (e). Therefore, no improvement could be reached using curve fitting. The next step is to use the variables' transformation with the aim to improve the MRE.

g) MRE with Transformations of Variables (ED, no Constant)

Another method that could improve the MRE is to transform a variable in order to generate a normal distribution. Various combinations have been tested and the combination with a natural logarithmic transformation for the variable "Surface" (Appendix 8.24) that reaches an adjusted R^2 of 0.929 is described. The following information summarises the criteria used to evaluate the MRE. The details about the model analysed are available in Appendix 8.23.

Table 5.31 g) MRE Regression Equation Coefficients with Significance

Model		Unstandardised Coefficients		t	Sig.
		B	Std. Error		
1	Lake_ha	.001	.000	7.783	.000
	Usage	.223	.032	6.872	.000
	DistanceToCentre_m	5.13E-005	.000	3.332	.001
	TR_LN_Surface_m2_Norm	.287	.041	7.017	.000

(Source: Developed from field data using SPSS)

The Table 5.31 indicates the regression equation coefficients with their significance for the four independent variables entered simultaneously into the model that uses the "EuclideanDist (ED)" as dependent variable and excludes the constant variable. The R^2 and adjusted R^2 are respectively 0.966 and 0.929. The ANOVA F-test is significant. Therefore the H_0 can be rejected proving a significant relationship between the dependent variable and the independent variables. The selected variables are "TR_LN_Surface," "Lake_ha," "Usage," and "DistanceToCentre." This model does not present evidence of particular problems. The resulting model presents a lower adjusted R^2 (0.929) than the model (e). Therefore, no improvement could be reached using variables' transformation method.

After the selection of possible candidates as predictor and after the various combinations of MRE are estimated and no substantial improvements are found against the model e); the model e) is chosen for a deeper examination of the multiple regression assumptions. Thus the model e) (Appendix 8.19) is tested against the regression assumptions in the following sub section.

5.9.2 Multiple Regression Assumptions' Analysis

Before applying any multivariate technique, the fit of the sample data is assessed with the statistical assumptions underlying that multivariate technique. For the regression analysis, the assumptions of normality, homoscedasticity, independence and linearity of residuals are assessed (Hair et al. 2006). Hair et al. (2006) affirm that a violation of the statistical assumptions may cause biases or non-significance in the results that cannot be distinguished from the true results. The assumptions have to be verified in order to increase reliability of the regression model.

Gujarati (2003) indicates the following twelve assumptions that should not be violated by the classical normal linear regression model in order to obtain significant and reliable results that can be used and interpreted: 1) Linear regression model, 2) x values are fixed in repeated sampling, 3) zero mean value of residuals, 4) homoscedasticity or equal variance of residuals, 5) no autocorrelation between the residuals, 6) zero covariance between residual and predictor or independence of residuals, 7) number of observations n must be greater than number of parameters to be estimated, 8) variability in predictors (x) values, 9) regression model is correctly specified, 10) there is no perfect multicollinearity, 11) the residuals (error terms) should be normally distributed and 12) multivariate outliers. *These twelve assumptions have been verified and the results do not show clear evidence of violation of them*. The details about the statistical values and graphical analyses are described in Appendix 8.19h.

5.10 Empirical Results and Conclusion

This chapter started with data analysis as the process of gathering, modelling, and transforming data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. After the data has been prepared, the PFs REP_benchmark has been computed and the "distance" as measurement between the PFs REP_benchmark and a specific PF's REP has been defined. The real estate portfolio empirical decision model (REP-EDM) has been developed with a multivariate regression in an iterative process in order to find the "best" regression model.

The REP-EDM generated by the selection of the "best" regression equation including the estimated coefficients (Appendix 8.19) can be summarised in the following multiple linear regression:

$$ED = (1.489E-02)x_1 + (5.667E-04)x_2 + (2.131E-01)x_3 + (4.622E-05)x_4 \qquad (5.1)$$

Where

- ED : Euclidean Distance represents the "distance" to REP_benchmark
- x_1 : represents the predictor "Surface_m2"
- x_2 : represents the predictor "Lake_ha"
- x_3 : represents the predictor "Usage" in percent
- x_4 : represents the predictor "DistanceToCentre_m"

The REP-EDM model is a representation of the PF RE market. The REP-EDM allows a PF manager to evaluate and compare its REP with a REP_benchmark that considers the physical characteristics of the PF RE market and without the need of RE specialists. The

PF manager can use this information as an additional decision support system to optimise its REP according to the strategy of its PF. The REP-EDM supports managers by presenting their REP situation transparently compared to the RE market and it contributes to extending the existing body of knowledge regarding REP management, transparency and understanding of the RE market in the Canton of Zürich. In brief, the aim of this chapter was to develop a model, based on the PFs REP_benchmark, which allows pension funds to analyse their REP for individual property types using an empirical model based on real physical characteristics of the REs owned by the PFs in Canton of Zurich.

6 DISCUSSION OF RESULTS AND CONCLUSIONS

6.1 Introduction

The purpose of this study is to develop an empirical decision model for real estate portfolio (REP-EDM) evaluation under consideration of real estate (RE) physical characteristics. The goal of the REP-EDM is to support PF managers to answer the research question of this study "*How can a customer's REP be optimised in order to reduce its idiosyncratic risks, basing the analysis on its physical characteristics and comparing it to a benchmark of the RE market physical characteristics?*" in a objective way without the need of RE specialists who base their knowledge only on their experience with local RE markets and without the introduction of a subjective human factor that influences the result.

The discussion in this last chapter is structured around the research question of this study presenting findings for the research issue within the context of prior research examined in the literature review. According to Perry (2002) and in order to preserve objectivity, Chapter Five was restricted to presentation and analysis of the collected data and to the development of the REP-EDM without drawing conclusions or comparing results to those of other researchers. Thus, the results have been separated from the discussion of their significance, interpretation, implications and contributions to the body of knowledge that is the purpose of this chapter.

In this chapter, the results are discussed, judged and interpreted, starting the analysis with the Section 6.2 containing the interpretation of the REP-EDM's variables and followed by the Section 6.3 containing two practical applications of the REP-EDM. The first application is the case study of pension fund Nr. 6 (PF6), which has been randomly selected and extracted from the PFs population during the development of the model (Subsection 5.6.4). The REP of PF6 is evaluated using the developed REP_benchmark and REP-EDM model and possible optimisation recommendations are made. The second practical application of the REP-EDM analyses the relationship between the individual PF sizes and their respective trading-related liquidity risks measured as "distance" to the REP_benchmark (liquidity risk) according to this study. The Section 6.4 presents the conclusion summarising the contributions of this study. The Section 6.5 discusses potential practical implications from this research for public and private sectors and presents an implementation procedure for managers. The Section 6.6 exhibits suggestions for future research.

6.2 Interpretation of the REP-EDM including the Predictors

In this correlational research, the strength of relationships among variables is measured so that the dependent variable can be predicted from the other independent variables. The multiple regression model is used to determine, through an interactive process, the statistical significance of each variable included or excluded to the final REP-EDM model. Statistical significance does not substitute the necessity of an interpretation of the results. The interpretation of the statistical selection of the most relevant variables included into REP-EDM is done with a practical significance analysis. In fact, whereas statistical significance determines whether the result is attributable to chance, practical significance assesses whether the result is useful in achieving the research objectives (Hair et al. 2006; Leedy & Ormrod 2005).

6.2.1 Model Fit - Coefficient of Determination

The chosen multiple regression model (Appendix 8.19) explains 93.3 percent (adjusted coefficient of determination R^2) of the ED variance. Although this value seems to be high, it does not explain 100 percent of the ED variance and there are various possible explanations for this gap. One reason for this gap could be that, due to limited data availability, not all existing RE characteristics have been included into the development of the model as independent variables, such as the condition of the building (new, rehabilitated, good maintained, to rehabilitate), quality of the location, quality of design and architecture, aesthetical dimensions and type of municipality. Again due to limited data availability the economic factors influencing the RE market are not considered for this study. For example, the current global and local markets influenced by economic cycles of expansion, prosperity, contraction and recession, legal constraints including environmental protection issues such as new standards for energy saving and sustainability in the construction of RE (Swiss Minergie Certificate), buyer or seller characteristics, RE trends, demography of population and RE brokers activities are not included in this analysis.

A further reason could be the desire for parsimony, avoiding an over-fitting of the model, including variables with an insubstantial contribution to the result as indicated by the adjusted R^2 which considers the number of independent variables in relation to the sample size. Still another reason could be the prevalent prediction goal of this model, that having a need for reliability strongly argues for a 'small maximum model' (Kleinbaum et al. 1998, p. 389). In brief, with an explanation of 93.3 percent of the variance, the proposed regression analysis is deemed sufficient to identify not only statistical significantly relationships but also relationships that have managerial significance.

6.2.2 Excluded Variables from REP-EDM

Four of the 17 independent variables (Appendix 8.1) considered in the model development have been included in the REP-EDM model which measures the liquidity risk of a PF REP by assessing its "distance" from the REP_benchmark. Before analysing these four relevant variables (Surface, Lake, Usage and DistanceToCentre), the 13 omitted physical characteristics are discussed. Although these variables have been excluded from the liquidity risk measurement model, they maintain a practical usability as indicators in the REP_benchmark in order to gain insights into the PF RE market, providing possible directions for optimisations of a PF REP.

<u>Structural variables</u>

1- Usage	Included in the REP-EDM and discussed in the next subsection.
2- Age	According to the literature about empirical studies, price modelling such as the hedonic price evaluation model, the variable "Age" is a significant variable (Studenmund 2006; ZKB 2004; Geltner & Miller 2001; Linneman 1980; Grether & Mieszkowski 1974). In the PFs-ZH-Dataset the variable "Age"

	shows a practical, independent relationship with the dependent variable "ED" with a non-significant correlation of 0.08 (Figure 5.8, Table 5.24). Therefore, it does not affect the model significantly.
3- NrRooms	Is omitted by the stepwise regression procedure due to the highly significant ($p < 0.01$) Person's correlation of 0.813 with the variable "Surface" (Table 5.15, Figure 5.9). In fact, the metrage included in the number of rooms of a RE does not add a substantial contribution that justifies an inclusion in the model, because the variable "Surface" indirectly represents this variable.
4- Surface	Included in the REP-EDM and discussed in the next subsection.
5- Volume	For the same reason as for the variable "NrRooms", it is omitted.

<u>Macro Variables</u>

6- GeoLocationX	The geographical coordinate can be considered implicitly included into the variable "DistanceToCentre" which is part of the model. For this reason reducing the variable "GeoLocationX" has been omitted.
7- GeoLocationY	For the same reason as for the variable "GeoLocationX", it is omitted.
8- DistanceToCentre	Included in the REP-EDM and discussed in the next subsection.
9- TaxationLevel	This variable does not have a contribution substantial enough to be included. One reason could be that, in the analysed REs-ZH-dataset, the mean of 123 percent with a standard deviation of 13

percent seems (Annex 8.5.9) to be a high value for most of the REs included into the REs-ZH-Dataset. The changes between these high values do not reach a sufficient contribution to be included into the model.

<u>Micro Variables</u>

10- Lake	Included in the REP-EDM and discussed in the next subsection.
11- View	This variable can be considered as implicitly included into the variables "DistanceToCentre" and "Lake" that are part of the model and this is an argument for its exclusion.
12- Slope	A possible argument for this variable's exclusion is that this variable has practically same value throughout all the REs analysed, with a mean of 3 percent and a standard deviation of 2.8 percent (Annex 8.5.12, Annex 8.6-12).
13- SunJuly	For the same reason as the variable "Slope", it has been omitted by the regression procedure. In fact, this variable has a mean value of 5,560 KJ/m^2 and a standard deviation of 140 KJ/m^2.
14- Dist_Bus	*)
15- Dist_Railway	*)
16- Dist_School	*)
17- Walk_Index	*)

*) These four variables do not show a significant contribution to the REP-EDM (Annex 8.15); therefore, they are omitted from the model. A possible interpretation of this result is the finding that the Canton Zurich has a high density network of public transportation

that allows a very good attainability of each place in the canton. The majority of the REs analysed have a bus station within 500 metres and a rail station within 1,000 metres (Annex 8.6-14, -15). With regard to schools and points of interest, the same reason as for the distances can be used. This high "attainability" is confirmed by the Figure 5.8 and by the histogram in the Annex 8.6-16, which shows a high density of schools present and that practically all the REs analysed have a school within 500 metres. The fact, that there is a small number of REs that are not easily reachable and the high homogeneity in attainability within the Canton Zurich can be the reasons for difficulty in finding a relevant influence of these variables on the model. In brief, the Canton Zurich is so easily reachable, that it is almost impossible to demonstrate a significant and relevant change in the risk value measured with the "distance" due to these variables.

6.2.3 Included Variables into REP-EDM

After the coefficients of the model are estimated, the regression variate is specified and the diagnostic tests are administered. The examination of the predictive equation based on the four selected independent variables can then begin. With an acceptable level of model fit of 93.3 percent, interpreting the variate reveals the nature of the multivariate relationship. The interpretation of the effects for individual variables is made by examining the estimated coefficients or weights for each of the four significant variables selected in the variate of the REP-EDM model (Surface, Lake, Usage and DistanceToCentre). The objective is to identify empirical evidence of multivariate relationships in the sample data analysed.

According to Graziano and Raulin (2000), the interpretation of a correlation starts by noting its direction in the sign and size in terms of contribution to the model. The estimated regression coefficients, termed the *b* coefficients, represent both the type of relationship (positive or negative) and the strength of the relationship between independent and dependent variables in the regression variate. This relationship indicates the change in the dependent value each time the independent variable changes by one unit. The regression coefficients play two key functions in meeting the objectives of prediction and explanation for the regression analysis (Hair et al. 2006). In the following, the coefficients are analysed according to these two key functions.

The four predictors “Surface“, “Lake“, “Usage“ and “DistanceToCentre” significantly contribute to the prediction of the Euclidean distance and all have a positive coefficient influencing the Euclidean distance (ED) positively. The interpretation of this result suggests that an increase of one of these variables increases the ED, respectively the liquidity risk in a PF REP. This finding confirms the definition of “distance” in this study. In fact, the consequence of an increase in the “distance” to the REP_benchmark corresponds to taking more risks into the REP. For example, buying a RE whose “DistanceToCentre” is higher than the same variable on the REP_benchmark would increase the risk of this REP.

Are these four variables relevant and significant in others’ studies too? The REP-EDM model includes the four predictors for the determination of the “distance” to the REP_benchmark. The importance and relevance of these variables for the characterisation of a RE is consistent among recent empirical studies reported in the literature. Although they followed other goals, such as the RE price estimation and the risk/return determination, they base their models on a common denominator including these four variables (Studenmund 2006; Prioni & Bignasca 2005; Häussermann et al. 2004; ZKB 2004; Geltner & Miller 2001; Linneman 1980; Rosen 1974; Grether & Mieszkowski 1974). In brief, the result of this study is supported by the literature.

The next question could be "which is the most important predictor in terms of influence on the liquidity risk in a REP in the REP-EDM model?" The size of the variables' coefficients depends on the units used for the measurement of the variables themselves. Therefore the unstandardised coefficients cannot be compared with each other. As such, the answer to the question can be extracted by analysing the standardised beta coefficients described in the Appendix 8.19. According to these coefficients, the major contribution to the model is given by the variable "Surface" with a beta of 0.489. The next predictor, in order of contribution, is the variable "Lake" with a beta of 0.306, the next is the variable "Usage" with a beta of 0.223 and the last is the variable "DistanceToCentre" with a beta of 0.196. A confirmation of the ranking of these variables can be found in the empirical pricing model developed by the ZKB (2004), in which one of the most relevant structural variables is the surface, one of the most relevant macro variables is the distance to centre and one of the most relevant micro variables is the lake view (STA 2008b; Prioni & Bignasca 2005).

To the extent that "Surface" could be changed, without impacting other variables, it represents the most effective way (higher beta), ceterus paribus, of influencing the risk of a REP. Of course, changing "Surface" is theoretically doable, but practically it is difficult to simply change the value of a physical characteristic without buying or selling an entire RE. With an unstandardised coefficient of 0.014 and an increase of one square metre of surface, a change of the Euclidean distance (ED) by 1.4 percent can be expected.

Although the variable "Usage" practically seems to have an insubstantial relationship of about 0.7 percent between SFD and CO in the REs-ZH-Dataset analysed (Annex 8.5.1), it is significant in the REP-EDM model. A possible explanation can be the diversification effect that is introduced into the REP by adding a SFD. A change of the ED by 0.05 percent is expected when increasing the visible "Lake" of a REP by one hectare. The importance of the variables "Lake" and "DistanceToCentre" is demonstrated with the graphical disposition of the REs on Figure 5.7 which seems to confirm the attraction of the two centres (Zurich and Winterthur) and the importance of a location near to a lake. The next consideration is that a REP can be evaluated and optimised and it is not only a question of a single variable and a single RE, but of the entire REP. This fact is supported by the research question of this study that relates the evaluation and optimisation of a REP with the REP_benchmark. A practical application which uses the REP-EDM model and the REP_benchmark to answer this question is presented in the next section.

6.3 Practical Applications of the REP-EDM

6.3.1 Out of Sample Analysis with the PF REP Nr. 6 (PF6)

The REP of the PF Nr. 6 (PF6) has not been used to develop the REP-EDM model, therefore an out of sample analysis is done with the goal of demonstrating how a PF manager can use the developed REP-EDM model and the multidimensional

REP_benchmark to evaluate and optimise its portfolio. This practical case study, which has been randomly selected (Section 5.6) from the PFs population, is analysed with the following four steps. First, a descriptive analysis is presented. Second, an evaluation using the developed REP-EDM model and the REP_benchmark is done. Third, possible optimisation recommendations under the assumption that the PF manager is risk-averse are suggested. At fourth, the implementation of the recommendations is simulated through the introduction of a virtual RE into the PF6 REP and its implications on the REP risk are analysed.

For the first step, some of the PF6 REP physical characteristics are depicted in Table 6.1 (because of data protection no more variables are presented). The PF6 presents a REP containing 19 REs, all condominiums, a surface pro RE ranging from 65 to 106 square metres, three different distances to centre indicating that the 19 REs are divided into three multi family dwelling as confirmed by the three different geographical coordinates, two taxation levels for the three MFD ranging from 122.45 to 129.53 percent and two surfaces of lake seen, the first of 1,635 hectares and the second of 5,442 hectares, indicating that two of three MFD lie close together and in the same municipality.

Table 6.1 REP Physical Characteristics of PF Nr. 6

PF Nr.	RE Nr.	Usage	Age	NrRooms	Surface m2	DistanceToCentre m	TaxationLevel Percent	Lake ha	...
6	1	CO	52	1	66.0	1'805.90	129.53	1'635.00	...
6	2	CO	52	3	72.0	1'805.90	129.53	1'635.00	...
6	3	CO	52	3	72.0	1'805.90	129.53	1'635.00	...
6	4	CO	52	4	84.0	1'805.90	129.53	1'635.00	...
6	5	CO	52	4	85.0	1'805.90	129.53	1'635.00	...
6	6	CO	52	4	92.0	1'805.90	129.53	1'635.00	...
6	7	CO	52	4	95.0	1'805.90	129.53	1'635.00	...
6	8	CO	52	5	106.0	1'805.90	129.53	1'635.00	...
6	9	CO	52	5	106.0	1'805.90	129.53	1'635.00	...
6	10	CO	45	3	65.0	18'219.08	122.45	5'442.00	...
6	11	CO	45	3	65.0	18'219.08	122.45	5'442.00	...
6	12	CO	45	3	65.0	18'219.08	122.45	5'442.00	...
6	13	CO	45	4	75.0	18'219.08	122.45	5'442.00	...
6	14	CO	45	4	75.0	18'219.08	122.45	5'442.00	...
6	15	CO	67	3	65.0	18'232.11	122.45	5'442.00	...
6	16	CO	67	3	65.0	18'232.11	122.45	5'442.00	...
6	17	CO	67	3	65.0	18'232.11	122.45	5'442.00	...
6	18	CO	67	3	65.0	18'232.11	122.45	5'442.00	...
6	19	CO	67	3	65.0	18'232.11	122.45	5'442.00	...

(Source: Developed for this research)

The second step starts by presenting in Table 6.2, the data used to evaluate the PF6 REP using the developed REP-EDM model and the PF's REP_benchmark. The means of the variables for each PF (μ_{iB}) and the coefficients of variation (cv_{iB}) are used for a direct REP comparative analysis with the variables μ_{i6} and cv_{i6} of the PF6. The former (μ_{iB}) gives an indication of the absolute values of the RE market and the latter (cv_{iB}) is used as relative measure of the scatter in the data relative to the mean of the variable.

The variable "Usage," with a mean and a *cv* of zero, confirms the result of the descriptive analysis that only CO are present in the PF REP Nr. 6 and presents a lower diversification in comparison with the REP_benchmark that contains 0.99 percent SFD. The variable "Age" identifies a 27.9 percent older REP for the PF6 compared to the REP_benchmark with a 47.9 percent lower *cv*. The variables "NrRooms, GeoLocationX, GeoLocationY, DistanceToCentre, TaxationLevel, View, Slope, SunJuly" present a gap lower than 10 percent in comparison to the REP_benchmark indicating a similarity to it. The variable "Surface," with a mean of 11.7 percent lower than the REP_benchmark, and the variable "Volume," with a mean of 38.0 percent lower than the REP_benchmark, indicate a tendency of the PF6 to own smaller RE with smaller rooms than the buildings of the other PFs in the Canton Zurich.

Table 6.2 REP_benchmark and Euclidean Distance with REP-EDM of PF Nr. 6

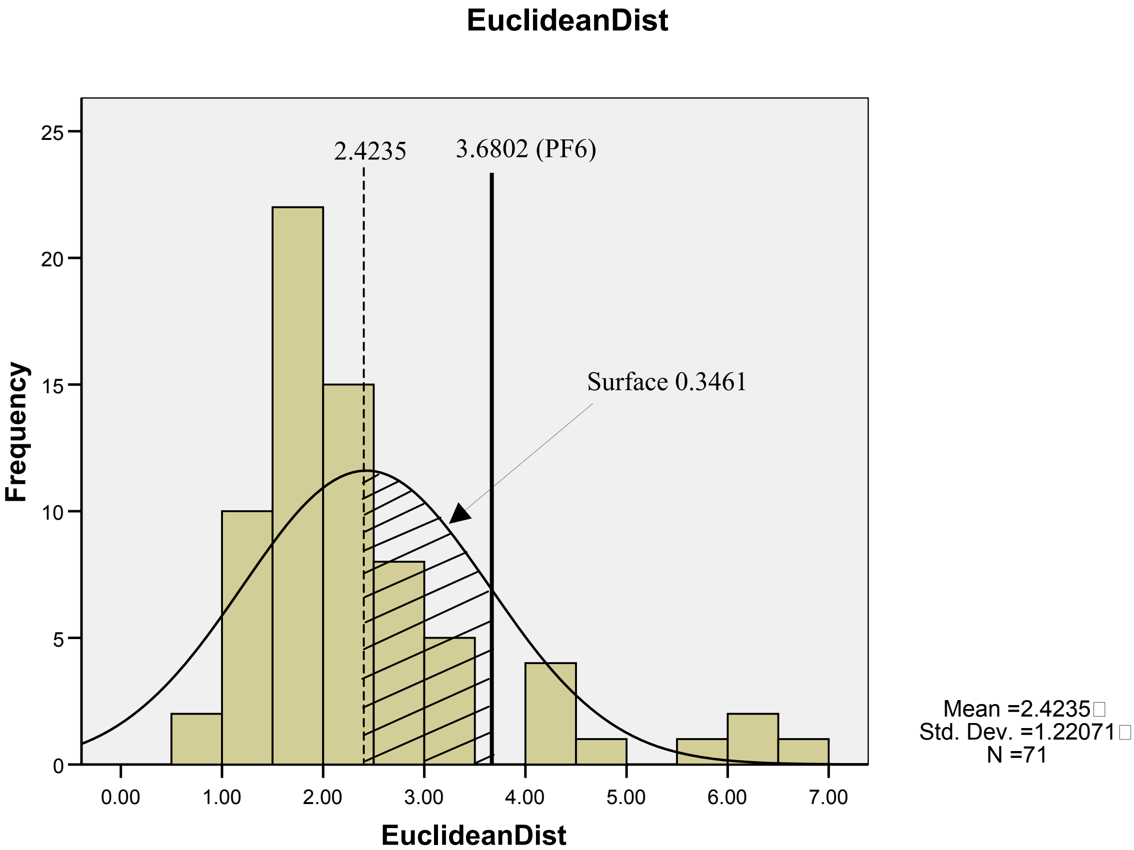
	REP_benchmark values between all PFs				PF Nr. 6	
Predictors	N	$NormV_{iB}$	$Mean_{iB}$	cv_{iB}	$Mean_{i6}$	cv_{i6}
Usage	71	1.00	0.99	2.69	0.00	0.00
Age	71	1.00	42.29	0.30	54.11	0.16
NrRooms	71	1.00	3.36	0.29	3.42	0.26
Surface_m2	71	1.00	86.37	0.27	76.21	0.19
Volume_m3	71	1.00	5420.78	0.34	3360.53	0.17
GeoLocationX_m	71	1.00	687531.31	0.00	688793.58	0.01
GeoLocationY_m	71	1.00	248825.17	0.01	238392.79	0.03
DistanceToCentre_m	71	1.00	9929.45	0.24	10447.84	0.81
TaxationLevel_Percent	71	1.00	121.63	0.04	125.80	0.03
Lake_ha	71	1.00	792.18	0.77	3638.68	0.54
View_ha	71	1.00	16068.83	0.31	16415.11	0.13
Slope_Percent	71	1.00	3.59	0.41	3.82	0.51
SunJuly_KJ_m2	71	1.00	5580.86	0.01	5478.21	0.00
Dist_Bus_m	71	1.00	170.69	0.30	90.10	0.49
Dist_Railway_m	71	1.00	884.15	0.26	732.45	0.27
Dist_School_m	71	1.00	243.91	0.34	96.25	0.32
Walk_Index	71	1.00	76.70	0.11	91.17	0.06

(Source: Developed for this research)

The variable "Lake," with a 359.3 percent higher mean and a 30.3 percent lower *cv,* indicates that the PF6 REP contains REs with a high surface of visible hectares of lake seen from the buildings, therefore it is possible that expensive RE are present in its REP (ZKB 2004). The variables "Dist_Bus," "Dist_Railway," "Dist_School" and

“Walk_Index” show that the PF6 REP contains RE that are situated nearer to the public transportations, schools and points of interest than the REs contained in the REP_benchmark. The next observation concerns the *cv* of the PF6 REP. In fact, 12 of 17 variables indicate a lower *cv* than the REP_benchmark indicating a low variation or diversification in type of RE contained in the PF6 REP. This is confirmed by the low number of REs (19) included in the PF6 REP compared to the mean of 222.48 on the REP_benchmark (Table 5.23).

Figure 6.1 Histogram Dependent Variable “EuclideanDist”



(Source: Developed for this research using SPSS)

The REP-EDM model with the Formula 5.1 is used to calculate the Euclidean distance (ED) for the PF6 and the result is the value of 3.6802 (Figure 6.1), which indicates the level of liquidity risk that the PF6 is taking in its REP in comparison to the PF population of the Canton Zurich. As depicted in the Figure 6.1, with the ED distribution that shows a mean of 2.4235 and standard deviation of 1.2207, it is possible to calculate the percentage of PFs in Canton Zurich that are taking more risks than the PF6. By standardising the ED distribution values (Appendix 8.28) and using the cumulative standardised normal distribution, a surface of 0.3461 (one sided) is calculated. The ED distribution of the PFs is two sided, therefore the result indicates that 69.22 percent of the PFs have a shorter "distance" than the PF6. On the other side, this means that 30.78 percent of the entire PFs population take more risks than the PF6 in terms of replicating the REP_benchmark.

Before starting with the third step of this practical analysis, suggesting optimisation recommendations for the PF6 management, the results of the evaluated PF6 REP are summarised. Which tendencies or physical characteristics are present in the PF6 REP compared to the REP_benchmark? The tendencies or physical characteristics present are: a) No SFD, b) older RE, c) smaller volumes and rooms, d) possible expensive RE, e) RE locations near to centre, bus, railway, schools and points of interest similar to benchmark, f) low variation or diversification in types of RE, g) PF6 REP includes 19 REs and h) Euclidean Distance of 3.6802 with 69.22 percent of the PFs taking less risks.

According to the REP-EDM evaluation, the REP_benchmark analysis and under the assumption that the PF manager is risk-averse, the following optimisation recommendations can be proposed in order to reduce the distance respectively the risk for the PF6. In form of a question: What should be changed in the REP of the PF6 in order to reduce its risk and thus replicate the REP_benchmark? Possible changes are: a) The PF6 seems to own a small portfolio (19 REs) compared to the REP_benchmark mean per PF (222 REs), therefore buying instead of selling REs is suggested, b) including SFD would increase the diversification in usage of the portfolio, c) buying newer RE (for example under 42 years of age) in order to reduce the average age in the PF6 REP, d) buying RE with more surface in terms of more surface per room and more volume per RE, e) buying RE with a lower surface of visible hectares of lake seen and f) buying RE with a distance to centre lower than 9,929 metres.

The final step of this practical case study includes the simulation of one additional virtual RE-object introduced into the PF6 REP using to the optimisation recommendations developed in the third step with the same assumption of a risk-averse management. Using these suggestions, different RE-objects could be found on the RE market and for this case study the following RE-objects (SFD, MFD 1 and MFD 2) depicted in Table 6.3 are assumed to be available on the market for this simulation. The problem to solve is to find out which one of these RE-objects can most clearly reduce the risk of the PF6 REP.

Table 6.3 Virtual REs on the Market

Predictors	SFD	MFD 1					MFD 2			
Usage	1	0	0	0	0	0	0	0	0	0
Age	2	10	10	10	10	10	22	22	22	22
NrRooms	5	3	3	4	4	5	4	4	4	4
Surface_m2	140	85	85	98	98	120	92	92	92	92
Volume_m3	910	4910	4910	4910	4910	4910	3929	3929	3929	3929
DistanceToCentre_m	5400	9030	9030	9030	9030	9030	6020	6020	6020	6020
TaxationLevel_Percent	115	120	120	120	120	120	121	121	121	121
Lake_ha	0	200	200	200	200	200	0	0	0	0
View_ha	15400	13500	13500	13500	13500	13500	14010	14010	14010	14010
Slope_Percent	3.10	2.80	2.80	2.80	2.80	2.80	3	3	3	3
SunJuly_KJ_m2	5600	5700	5700	5700	5700	5700	5100	5100	5100	5100
Dist_Bus_m	440	520	520	520	520	520	602	602	602	602
Dist_Railway_m	1200	1000	1000	1000	1000	1000	980	980	980	980
Dist_School_m	780	2200	2200	2200	2200	2200	1450	1450	1450	1450
Walk_Index	71	72	72	72	72	72	81	81	81	81

(Source: Developed for this research)

The results of separately introducing all three virtual RE-objects (SFD, MFD 1, MFD 2) into the REP-EDM model are presented in Table 6.4.

Table 6.4 Virtual REs Simulation with REP-EDM

	PF Nr. 6		PF Nr. 6 + SFD		PF Nr. 6 + MFD 1		PF Nr. 6 + MFD 2	
Predictors	*Mean*	*cv*	*Mean*	*cv*	*Mean*	*cv*	*Mean*	*cv*
Usage	0.00	0.00	0.05	4.47	0.00	0.00	0.00	0.00
Age	54.11	0.16	51.50	0.28	44.92	0.44	48.52	0.30
NrRooms	3.42	0.26	3.50	0.27	3.50	0.25	3.52	0.24
Surface_m2	76.21	0.19	79.40	0.25	80.58	0.20	78.96	0.18
Volume_m3	3360.53	0.17	3238.00	0.24	3683.33	0.22	3459.39	0.16
DistanceToCentre_m	10447.84	0.81	10195.45	0.81	10152.46	0.74	9677.78	0.81
TaxationLevel_Percent	125.80	0.03	125.26	0.03	124.59	0.03	124.97	0.03
Lake_ha	3638.68	0.54	3456.75	0.60	2922.29	0.77	3005.87	0.75
View_ha	16415.11	0.13	16364.35	0.13	15807.79	0.14	15996.83	0.14
Slope_Percent	3.82	0.51	3.78	0.51	3.61	0.49	3.68	0.49
SunJuly_KJ_m2	5478.21	0.00	5484.30	0.01	5524.42	0.02	5412.44	0.03
Dist_Bus_m	90.10	0.49	107.59	0.83	179.66	1.02	179.12	1.13
Dist_Railway_m	732.45	0.27	755.82	0.29	788.19	0.27	775.50	0.26
Dist_School_m	96.25	0.32	130.43	1.20	534.53	1.63	331.68	1.58
Walk_Index	91.17	0.06	90.16	0.08	87.18	0.11	89.40	0.07
Euclidean Distance	3.6802		3.6236		3.3257		3.3269	
PFs with higher ED	30.7%		32.7%		45.9%		45.8%	

(Source: Developed for this research)

All the three simulated RE-objects, which have been chosen according to the optimisation recommendations, improve the risk situation of the PF6 reducing the Euclidean distance. The introduction of the SFD reduces the distance by 1.54 percent, the MFD 1 by 9.63 percent and the MFD 2 by 9.60 percent. According to the REP-EDM model, the RE-object that a risk-averse manager would choose is the MFD 1, which includes five REs. In reality, other factors may influence the decision of a PF manager and in this case study, a possible criterion could be the relative relationship between the reduced risk by the introduction of a MFD in the REP and the additional price of the MFD. Following this criterion, the same manager would choose the MFD 2 that includes only four REs, therefore probably cheaper than the MFD 1 (ceterus paribus). In fact, with about the same value of risk reduction, the additional price paid for the MFD 1, which includes five REs, cannot be justified. Another criterion could be the mode or the perceived quality living in a RE with lake view. In this case the choice would remain on the MFD 1, if also probably more expensive.

Not all managers are risk-averse and the following question can arise: It is better to be near to or far away from the REP_benchmark? To answer this question, first the PF strategy has to exist or be defined in which the risk-aversion level determined, then the REP-EDM model and REP_benchmark can be used to support the implementation of the PF strategy. The REP-EDM does not substitute managerial decisions regarding the PF strategy or which RE has to be bought, but it supports managers by presenting their REP situation transparently compared to the RE market. In brief, the REP-EDM is based only on physical characteristics and no additional component such as investment, pricing, costing or return strategy are considered. Therefore PF managers should use not the REP-EDM model and the REP_benchmark as sole basis for their decisions, but instead incorporate this model in arrange of information for their decision-making process.

6.3.2 PF Risk (Euclidean Distance) versus PF Size

This practical application of the REP-EDM analyses the relationship between PFs size and their trading-related liquidity risks testing the following hypothesis.

H_0: Risk and pension fund size are not correlated.

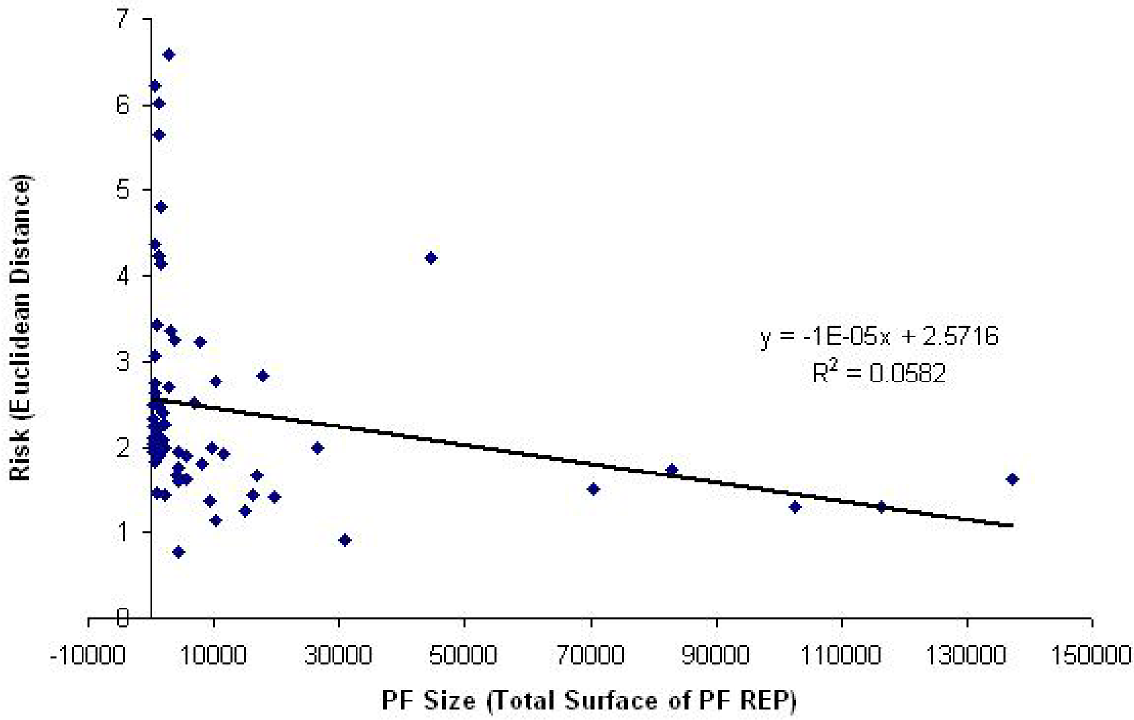
H_1: Larger pension funds take less risks, where risk is defined as the "distance" measure introduced in this research.

The risk carried by a PF is defined in this study as a function of the PF's "distance" to the PF REP_benchmark (Figure 1.4). This distance, the risk, is not a function of the size, therefore it is justifiable to hypothesise that risk and size are (or are not) correlated. The null hypothesis (H_0) is that there is no relationship between risk and size of the pension fund. With the application of the developed REP-EDM, it is possible to test this hypothesis calculating the risk for each PF independently.

Before starting with the analysis, the word "large" for a PF has to be defined in terms of physical RE characteristics. "Large" is a measure on a cardinal scale and it gives an indication about the size of a PF. The size of a PF can be defined using various measurement such as the number of total rooms in the PF REP, the number of objects owned by the PF, the total volume of the PF REP, the total surface contained into the PF REP, the total value of all objects owned by the PF and the sum of the PF's balance sheet.

According to the data availability, "Large" is defined in this practical application with the total surface contained into the PF REP. This measure not only indicates the size of a PF through a physical characteristic, but it gives, according to the hedonic price modelling, an indirect indication about the size in the value or total price of the REs contained into the PF REP (Studenmund 2006; Prioni & Bignasca 2005; Häussermann et al. 2004; ZKB 2004; Geltner & Miller 2001; Linneman 1980; Rosen 1974; Grether & Mieszkowski 1974). Additionally, this measure detects the difference between a REP containing three small (in surface) CO with another REP containing three big CO, that would not be identified using the total number of room or total number of REs owned by a PF.

Figure 6.2 Diagram PF Risk (Euclidean Distance) versus PF Size



(Source: Developed for this research)

Table 6.5 Correlation PF Risk (Euclidean Distance) versus PF Size

a) Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	SurfaceProPF	-	Enter

b) Coefficients

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.5716	.157		16.389	.000
	SurfaceProPF	-1.09E-05	.000	-.241	-2.050	.044

(Source: Developed for this research using SPSS)

The null hypothesis is tested using a correlational analysis and the results are depicted in Figure 6.2 and Table 6.5. Although the regression model explains 5.8 percent (coefficient of determination R^2) of the variance, the correlation between PF risk and the PF size of −1.09E-05 is significant ($p < 0.05$), therefore the null hypothesis can be rejected. The H_1 can be accepted, in fact, the negative sign indicates that the greater the PF size is, the lower the risk. A possible reason, confirmed by others empirical studies (Braun et al. 2008; Montezuma 2004; Amman & Scherer 2001), that a larger PF takes less risk than a smaller PF is the increased diversification effect that a larger PF can apply, due in part to the higher financial resources available to directly invest in REs and its probably higher risk tolerance.

6.4 Conclusion about Research Issue and Research Question

The objective of this study is reached with the developed REP-EDM that may be used to answer the research question of this study. The REP-EDM model does not substitute other REP optimisation models mentioned in the literature review but instead it represents an additional model that helps managers make decisions in a market that is characterised by low transparency and inefficiency. The issue of estimating liquidity risk is crucial in developing a successful REP strategy and the REP-EDM contributes by extending the existing body of knowledge regarding the REP management, the insights and understanding of the RE market. An overview of the contributions made by this study is presented in Table 6.6. Each contribution is identified and referenced to the literature and the contribution summarised and referenced to the appropriate sections in this study.

Table 6.6 Overview Contributions from this Research

A Decision Model for Real Estate Portfolio Valuation and Optimisation

Under consideration of real estate physical characteristics

a) PF RE market Transparency
The need of a higher efficient RE market with increased transparency.
Coverage in the Literature: To some extent
Sections 1.5, 3.2, 3.3, 3.4, 3.4.6 and 3.5
Contribution from this research:

- New benchmark "REP_benchmark" for REP analysis has been built (Sect. 5.4).
- New model "REP-EDM" for REP analysis has been built (Sect. 5.8.1, 6.2.3, 6.3.1).
- The developed REP-EDM (Sect. 5.8.1) based on a REP_benchmark (Sect. 5.4) increases RE market transparency allowing the positioning of a specific REP compared to the PF RE market (Sect. 6.3.1).
- Increased transparency indicating which are the most important physical characteristics for the liquidity risk of a PF REP (Sect. 5.8.1), when the risk is defined as the "distance" to the REP_benchmark (Sect. 5.5, 6.2.3).

b) PF RE market Understanding
The need of better understanding of the PF RE market.
Coverage in the Literature: To some extent
Sections 1.5, 3.2, 3.3, 3.4, 3.4.6 and 3.5
Contribution from this research:

- New benchmark "REP_benchmark" for REP analysis has been built (Sect. 5.4).
- New model "REP-EDM" for REP analysis has been built (Sect. 5.8.1, 6.2.3, 6.3.1).
- Increased understanding for the body of knowledge explaining which are the most important physical characteristics for the liquidity risk of a PF REP (Sect. 5.8.1) winning new and interesting insights into the function manner of the PF RE market (Sect. 6.2.3).
- Increased insights of the PF RE market through the results of the relationship analysis between the PF REP physical characteristics and the real empirical PF REP_benchmark (Sect. 5.4, 5.8.1, 6.3.1)
- Distinct contribution to the body of knowledge by using the MLR (multiple linear regression) methodology, that represents a standard for the pricing modelling in RE evaluation, in a new way for the liquidity risk analysis of PF REP in Canton Zurich. The MLR has not been used before as in this study and it is the first time that a such data analysis has been done for Switzerland PFs in Canton Zurich (Sect. 5.7).
- Increased understanding of the PF RE market using the REP-EDM for the evaluation and optimisation of a specific PF REP (Sect. 5.8.1, 6.3.1) and for the analysis of the relationship between the PF size and its risk (Sect. 6.3.2).
- Due to the transparency in the standardised RE measurements (Sect. 4.6) and the statistical methods used, the REP-EDM could be computed for other areas enhancing understanding of the RE market (Sect. 4.12.5, 6.3.1, 6.3.2).

c) Objective Analysis of REP for Strategic Decision
The need of higher objectivity in the evaluation of a REP and the need for greater certainty and precision about RE measurements of risk in supporting strategic decisions.
Coverage in the Literature: To some extent
Sections 1.3, 1.5, 3.2, 3.4 and 3.4.6
Contribution from this research:

- The developed REP-EDM built on a REP_benchmark is based only on RE physical characteristics without any human factors. This fact increases the objectivity of the result and the market transparency in the RE business supporting strategic decisions. Using the REP-EDM, decisions on new investments and optimisations of the allocated resources can be taken in a more objective way and the REP_benchmark allows to be used by practitioners as the basis for REP decision-making (Sect. 6.3).
- PF manager can use the REP_benchmark to perform a comparative analysis in order to increase understanding and transparency of its REP, thus enhancing and supporting strategic decisions with the aim to optimise the REP (Sect. 6.3.1, 6.3.2).
- The REP-EDM based on MLR can be used as a mechanical decision supporting system (Sect. 6.3.1, 6.3.2).

---

d) REP Valuation without being a RE Specialist
The need for RE investors in making quicker investment decisions without having to be an expert in the local RE market.
Coverage in the Literature: To a very small extent
Sections 1.3 and 3.2
Contribution from this research:

- This study offers a decision supporting system “REP-EDM” (Sect. 5.8.1) including the “REP_benchmark” (Sect. 5.4) that can be used without being a RE specialist and a manager can obtain important information about its REP in order to take decisions for an optimisation of its REP (Sect. 6.3.1, 6.3.2).

---

e) PF REP Trading Liquidity Risk Management
The need of a liquidity risk measure for a REP
Coverage in the Literature: To a very small extent
Sections 1.5 and 3.3
Contribution from this research:

- A significant research contribution of this study building a REP_benchmark (Sect. 5.4) into the PF area and providing the possibility to compare it with a specific REP through the REP-EDM model (Sect. 5.8.1), is to permit the determination of the liquidity risk level “distance” of the analysed PF REP compared to the benchmark of the RE market (Sect. 5.5). This allows a better management and the optimisation of the idiosyncratic risks. Increasing liquidity risk awareness and transparency on real estate market. Reducing the risk of not being able to sell/buy when needed for PFs that decide to situate their REP near to the REP_benchmark (Sect. 6.3.1, 6.3.2).

(Source: Developed for this research)

The Table 6.6 includes the following five contributions: a) PF RE market transparency, b) PF RE market understanding, c) objective analysis of REP for strategic decision, d) REP evaluation without being a RE specialist and e) PF REP trading liquidity risk management. In summary, the contribution of the developed REP-EDM model, which includes the REP_benchmark, is to increase transparency and understanding of the PF REP market, enhancing the decision-making process with an objective and reliable REP evaluation model that includes estimations of the REP risks. This REP evaluation model represents an auxiliary decision support system for investors that can be used to evaluate and optimise a REP additionally to or independently from the RE expert opinions.

6.5 Implications for Policy and Practice

6.5.1 Private Sector Manager

In today's business world, the need for continuous improvement to beat competition is a constant and an increased PF REP market understanding and transparency have the potential to better equip PF managers so that they can optimise their REP according to their strategies and risk-aversion levels, thus gaining a competitive advantage by using benchmarking as an instrument for competitive analysis and performance evaluation. PFs can use the PF REP_benchmark to perform a comparative analysis in order to increase understanding and transparency of their REP, thus supporting strategic decisions. With a more efficient PF RE market there is also the potential to move more investors to investing in direct RE, increasing their fix assets, the sell and buy transactions and the total traded RE volume on the RE market while increasing the liquidity on the RE market. Two practical examples are presented in the Subsections 6.3.1 and 6.3.2.

The REP-EDM model has been presented for inspection to three CEOs of major corporations in the area under investigation. The first corporation is a real estate data broker, the second corporation buys and manages real estate and the third is a pension fund. They have expressed an appreciation for the focus of this research and the resulting increase in transparency on the RE market. They appreciate the opportunity to consider the REP-EDM and the REP_benchmark as an additional supporting tool in understanding the PF RE market and as a further connection between the academic and applied business research.

The REP-EDM has the potential to be applied by PF portfolio managers and investors as this new model allows for an alternative method to objectively evaluate a PR REP and to determine the change in liquidity risk for a new investment compared to the use of traditional portfolio and investment analysis. The practical significance is presented in the Subsection 6.3.1, in which the benchmarking model developed in this study is used both as an instrument to evaluate a PF REP and as an instrument for decision-making in the REP optimisation process. In brief, for the PF portfolio manager, the REP-EDM can be used as an additional decision supporting system for the management of its REP. For the investor, this benchmarking model can be used to analyse and compare possible single direct RE investments without being a local RE specialist. Thus, the benchmarking model developed in this study has clear practical applications for PF managers and investors in the Canton Zurich.

6.5.2 Public Sector Policy Analysts and Managers

The PFs have the exclusive purpose of financing pension plan benefits and they are regulated by Swiss law which defines the rules for their investments including direct RE investments. This study, delivering greater transparency of the PF RE market with practical guidance for using the REP-EDM and REP_benchmark has the potential to influence the regulations regarding asset categories and their maximum authorised weightings. Specifically, the category of direct RE investment could be adjusted according to the risk aversion of a PF. In fact, the policy maker can evaluate the risks taken by a PF with the additional support of the REP-EDM model. The weights of the direct RE category and the PF reserves (security margin) could be adapted according to the risk situation of the single PF with the goal of increasing security for the fulfilment of the designated payout of benefits.

6.5.3 Recommended Course of Action in the Practical REP Analysis

A goal of benchmarking in PF REP is to enable a PF to check and evaluate its situation compared to the other players on the RE market, as represented by the REP_benchmark. The following checklist presents a possible simplified procedure for PF managers who can incorporate the PF REP benchmarking model in their companies in order to obtain indications that support strategic decisions. For managers that wish to conduct a benchmarking process in their firms, the Xerox benchmarking model presented in Figure 3.2 can be used as basis and this checklist can supply additional steps focused on the PF REP analysis based on the findings obtained in this study.

Checklist for practical use of REP-EDM and REP_benchmark

1) Set the RE strategy for the PF according to the overall PF strategy, the PF risk aversion and the PF affordable risk. A question to answer is "should the PF REP replicate the REP_benchmark reducing its idiosyncratic risks or should the PF REP follow another focus such as including most of luxury and special REs?"
2) Calculate means and coefficients of variation of the current PF REP for the variables listed in Appendix 8.1 as done in Tables 5.21 and for the practical application in Table 6.2.
3) Use the REP_benchmark to compare and analyse the differences with the calculated parameters under number 2, to identify the direction of the optimisation. This should to be done in according to the defined PF strategy. As example, for risk-averse PF managers and for PF that cannot afford high risks, the course of actions should be defined with the goal of replicating the REP_benchmark (Table 5.21).
4) Use the REP-EDM model to identify which are the best REs to invest in or disinvest. An example of a practical application is presented in the Subsection 6.3.1.
5) Use the information and the results elaborated under points 3 and 4 as additional support for the strategic decision regarding whether to buy or sell REs and which type of REs should be transacted in order to fulfil the PF RE strategy.

6.6 Directions for Future Research

During the entire research process starting with the first analysis of the research question, the literature review, the development of the methodology, the data collection and analysis done with the aim of meeting the research objective, various directions for future research have been identified. In the following, possible future research which could further enhance the RE market understanding and transparency are suggested.

1) The first could be to use the same principle of the benchmarking model developed in this study and enlarging it fundamentally to other areas such as banking, assurance and to all institutional investors or to wealthy customers that have a REP.
2) The second could be to extend this study with a longitudinal or time series analysis. In this case, the same type of information would be collected across multiple time periods and a comparison analysis such as example among different years could be done. In addition to gaining insight into the RE market, for example with regard to the impact of the global financial crisis on the real estate market, another argument for this potential study extension is the continuous change of the population's needs over time. For example, an increase in the required surface per room over the years up to date.
3) The third as extension of the first two possible future researches could be to build a specific REP_benchmark index. The index could be based only on physical REP characteristics of the RE market and this could be done not only for PF but also for other areas.
4) Another extension of this research could be to analyse the relationship between the results of the REP analysis based on the REP-EDM of this study with other REP analyses that start from other basis, such as return on investment and risks or pricing and costs.
5) A geographical extension could offer another possible research project. Although the results of this study are valid only for the Canton Zurich, the development of the same

model for other regions, for the whole Switzerland and in other countries could be done.

6) A further extension could be to increase the number of variables in the study. The number of variables included into the developed model is restricted due to limitations on the data available. Thus not all existing or possible RE characteristics have been integrated. Depending on their availability, further independent quantitative and qualitative variables such as the condition of the building, quality of the location, quality of design, economic factors, legal constraints, buyer or seller characteristics, land area, RE trends, demography of population and environmental protection parameters could be included.

7 LIST OF REFERENCES

Alstete, JW 2008, ‘Measurement benchmarks or “real” benchmarking?: An examination of current perspectives’, *Benchmarking: An International Journal*, vol. 15 no. 2, pp. 178-186, viewed 19 September 2008, <www.emeraldinsight.com/1463-5771.htm>.

Anand, G & Kodali, R 2008, ‘Benchmarking the benchmarking models’, *Benchmarking: An International Journal*, vol. 15 no. 3, pp. 257-291, viewed 19 September 2008, <www.emeraldinsight.com/1463-5771.htm>.

Amman, D & Scherer, H 2001, *Immobilien als Baustein der Vermögensanlagen*, PPCmetrics AG, Zurich, Switzerland, viewed 24 April 2007, <www.ppcmetrics.ch>.

Apgar, M 1995, ‘Managing Real Estate to Build Value’, *Harvard Business Review*, vol. 73 issue 6, November, pp. 162-173, viewed 20 July 2007, EBSCOhost database, AN 9512052742.

Archer, WR & Ling, DC 1997, ‘The three dimensions of real estate markets: linking space, capital and property markets’, *Journal of Real Estate Finance and Economics*, vol. 14, no. 3, pp. 7-14, viewed 24 August 2009, EBSCOhost database.

Benjamin, JD, Sirmans, GS & Zietz, EN 2001, ‘Returns and Risk on Real Estate and other Investments: More Evidence’, *Journal of Real Estate Portfolio Management*, vol. 7 issue 3, July, pp. 183-214, viewed 25 July 2007, EBSCOhost database, AN 6464478.

Braun, U, Hasenmaile, F, Kaufmann, P, Neff, M & Rieder, T 2008, *Swiss Issues Real Estate: Real Estate Market 2008, Facts and Trends*, Credit Suisse Economic Research, Zurich, Switzerland, viewed 31 August 2008, <https://entry4.credit-suisse.ch/csfs/research/p/d/en/publikationen/pub_immobilienstudi>.

Booth, PM & Marcato, G 2004, ‘The dependency between returns from direct real estate and returns from real estate shares’, *Journal of Property Investment & Finance*, vol. 22 no. 2, pp. 147-161, viewed 11 May 2008, <www.emeraldinsight.com/1463-578X.htm>.

Camp, RC 1989, *Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance*, ASQC Quality Press, Milwaukee, WI.

Chapman, FM 1979, ‘Optimal Real Estate Portfolios’, *Journal of the American Real Estate & Urban Economics Association*, vol. 7 issue 3, pp. 298-317, viewed 25

July 2007, EBSCOhost database, AN 5863155.

Coakes, SJ & Steed, LG 2007, *SPSS: Analysis without anguish using SPSS Version 14.0 for Windows,* John Wiley & Sons, Australia.

Coleman, MS & Mansour, A 2005, 'Real Estate in the Real World: Dealing with Non - Normality and Risk in an Asset Allocation Model', *Journal of Real Estate Portfolio Management*, vol. 11 issue 1, April, pp. 37-53, viewed 26 July 2007, EBSCOhost database, AN 15794032.

Collins, TR, Rossetti, MD, Nachtmann, HL & Oldham, JR 2006, 'The use of multi-attribute utility theory to determine the overall best-in-class performer in a benchmarking study', *Benchmarking: An International Journal*, vol. 13 no. 4, pp. 431-446, viewed 11 May 2008, <www.emeraldinsight.com>.

Connor, G & Korajczyk, RA 2009, 'Factor Models in Portfolio and Asset Pricing Theory', viewed 24 August 2009, <http://ssrn.com/abstract=1139062>.

Coolican, H 1990, *Research Methods and Statistics in Psychology*, Hodder and Stoughton, London.

Crouhy, M, Galai, D & Mark, R 2001, *Risk Management*, McGraw-Hill Companies, Blacklick, OH, USA, viewed 15 October 2007, <http://site.ebrary.com/lib/unisouthernqld/Doc?id=5001801>.

CS - Credit Suisse Bank 2008, *Credit Suisse Swiss Pension Fund Index*, Credit Suisse Bank, Zurich, Switzerland, viewed 31 August 2008, <www.credit-suisse.com>.

Daniel, M 2008, *GVZ - Abgabe bzw. Verwendung von Daten der Gebäudeversicherung Kanton Zürich*, 8050 Zurich, Switzerland, email <info@gvz.ch>, confirmation letter of 1 December 2008.

Douglas, GA 1998, *The Management of Strategic Systems: The Quality Perspective*, 3rd edn, Pearson Custom Publishing, Boston, USA.

EFQM - European Foundation for Quality Management 2008, *EFQM: The Fundamental Concepts of Excellence*, Brussels, Belgium, viewed 2 May 2008, <http://www.efqm.org>.

Elmuti, D & Kathawala, Y 1997, 'An overview of benchmarking process: a tool for continuous improvement and competitive advantage', *Benchmarking for Quality Management & Technology*, vol. 4, no. 4, p. 229-246, viewed 10 May 2008, EBSCOhost database.

Encarta 2007, 'Benchmark', *World English Dictionary*, Bloomsbury Publishing Plc., viewed 31 August 2007, <http://encarta.msn.com/dictionary_/benchmark.html>.

Elton, EJ & Gruber, MJ 1998, 'Modern Portfolio Theory, 1950 to Date', viewed 24

August 2009, <http://ssrn.com/abstract=1297051>.

Falk, B 2000, *Fachlexikon Immobilienwirtschaft: Encyclopaedia of Real Estate Management*, Rudolf Müller, Cologne.

Fama, EF 1970, 'Efficient Capital Markets: A Review of Theory and Empirical Work', *The Journal of Finance*, vol. 25 issue 2, May, pp. 383-417, viewed 15 September 2007, EBSCOhost database, AN 4660197.

Fama, EF & French, KR 1996, 'Multifactor Explanations of Asset Pricing Anomalies', *The Journal of Finance*, vol. 51 issue 1, March, pp. 55-84, viewed 28 August 2007, EBSCOhost database, AN 960326250.

Fisher, JD & Geltner, D 2000, 'De-Lagging the NCREIF Index: Transaction Prices and Reverse-Engineering', *Journal of Real Estate Finance*, vol. 17 issue 1, pp. 7-16, viewed 18 October 2007, EBSCOhost database, AN 3083719.

Fong, PS, Shen, Q & Cheng, EWL 2001, 'A framework for benchmarking the value management process', *Benchmarking: An International Journal*, vol. 8 no. 4, pp. 306-316, viewed 19 September 2008, <www.emerald-library.com/ft>.

Freeman, AM 1993, *The Measurement of Environmental and Resource Values: Theory and Methods,* 2d edn, Resources for the Future, Washington D.C.

Fuhrer, J 2006, *Die berufliche Vorsorge in der Schweiz: Kennzahlen der Pensionskassenstatistik*, Bundesamt für Statistik (BFS), Neuchâtel, Switzerland, viewed 21 May 2007, <www.bfs.admin.ch>.

Garson, GD 2007, 'Testing of Assumptions', *Statnotes: Topics in Multivariate Analysis*, viewed 3 April 2007, <www2.chass.ncsu.edu/garson/pa765/assumpt.htm>

Geltner, DM & Miller, NG 2001, *Commercial Real Estate Analysis and Investment*, South-Western Publishing, Thomson Learning, Ohio, USA.

Georgiev, G, Gupta, B & Kunkel, T 2003, 'Benefits of Real Estate Investment', *Journal of Portfolio Management*, special issue 2003, pp. 28-33, viewed 2 September 2008, EBSCOhost database, AN 19923935.

Gleich, R, Motwani, J & Wald, A 2008, 'Process benchmarking: a new tool to improve the performance of overhead areas', *Benchmarking: An International Journal*, vol. 15 no. 3, pp. 242-256, viewed 19 September 2008, <www.emeraldinsight.com/1463-5771.htm>.

Graziano, AM & Raulin, ML 2000, *Research methods: a process of inquiry*, 4th edn, Allyn & Bacon, Needham heights, MA, pp. 174-98, cited in USQ 2005, *MGT8401 Research Methodology 1 – Selected readings 2*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 10.1.

Grether, GM & Mieszkowski, P 1974, 'Determinants of Real Estate Values', *Journal of Urban Economics*, pp. 127-146.

Guba, EG & Lincoln, YS 1994, 'Competing paradigms in qualitative research', in NK Denzin & YS Lincoln (eds), *Handbook of qualitative research*, Sage, Thousand Oaks, pp. 105-17.

Gujarati, DN 2003, *Basic Econometrics*, 4th edn, McGraw Hill, Boston, pp. 65-75, cited in USQ 2007, *MGT8402 Research Methodology 2 – Selected readings*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 4.2.

GVZ - Building Assurance Zurich 2008, Switzerland, <www.gvz.ch>.

Hair, JF, Black, WC, Babin, BJ, Anderson, RE & Tatham, RL 2006, *Multivariate data analysis*, 6th edn, Pearson Prentice Hall, Upper Saddle River, New Jersey.

Halbherr, P 2007, *Immobilienderivate*, Zürcher Kantonalbank, viewed 19 May 2007, <www.zkb.ch>.

Häussermann, H, Siebel, W & Wurtzbacher, J 2004, *Stadtsoziologie: Eine Einführung*, Campus Verlag, Zurich, Switzerland, viewed 31 August 2008, < http://books.google.ch/books?id=oVjcnjP9oCIC&pg=PA126 >.

Helbling, C & Leutwyler, O 2007, *Gesetzgebung der beruflichen Vorsorge*, Gemini Sammelstiftung, Zurich, Switzerland.

Hill, RC, Griffiths, WE & Judge, GG 2001, *Undergraduate Econometrics*, 2d edn, John Wiley & Sons, Australia.

Hiltmann, R 2007, CEO of the Pension Funds of the Zurich Cantonal Bank, Zurich, Switzerland, email <raymonde.hiltmann@zkb.ch>, interview of 29 August 2007.

Hurni, K & Stocker, PA 1996, *Diversifikation von Industrieportefeuilles unter Liquiditätsaspekten*, Bank- und finanzwirtschaftliche Forschungen, Universität St. Gallen, Switzerland.

IASB - International Accounting Standards Board 2005, *Element 1: Asset Definition*, International Accounting Standards Board, London, England viewed 17 March 2008, < www.iasb.org >.

IAZI - Informations- und Ausbildungszentrum für Immobilien AG 2007, Switzerland, viewed 15 October 2007, <http://www.iazi.ch/web/Indizes>.

Jain, R, Yadav, OP & Rathore, APS 2008, 'The propagation of benchmarking concepts in Indian manufacturing industry', *Benchmarking: An International Journal*, vol. 15 no. 1, pp. 101-117, viewed 19 September 2008, <www.emeraldinsight.com/1463-5771.htm>.

Kachigan, SK 1991, *Multivariate Statistical Analysis: A Conceptual Introduction*, 2nd edn, Radius Press, New York, USA.

Kerlinger, FN & Lee, HB 2000, *Foundations of behavioural research,*, 4th edn, Wadsworth, Thompson Learning, USA, pp. 557-73, cited in USQ 2006, *MGT8401 Research Methodology 1 – Selected readings 2*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 9.2.

Kleinbaum, DG, Kupper, LL, Muller & Nizam, A 1998, 'Selection the best regression equation', *Applied regression analysis and multivariable methods*, 3rd edn, Duxbury Press, California, pp.386-422, cited in USQ 2007a, *MGT8402 Research Methodology 2 – Selected readings*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 5.2.

Kumar, R 1996, *Research methodology: a step-by-step guide for beginners,* Longman, Melbourne, Australia, pp. 30-3, cited in USQ 2005, *MGT8401 Research Methodology 1 – Selected readings 1*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 4.1.

Leedy, PD & Ormrod, JE 2005, *Practical research: planning and design*, 8th edn, Merrill Prentice Hall, Upper Saddle River, New Jersey.

Levine, DM, Stephan, D, Krehbiel, TC & Berenson, ML 2005, *Statistics for Managers*, 4th edn, Prentice Hall, Upper Saddle River, New Jersey.

Lim, LC, McGreal, S & Webb, JR 2008, 'Institutional Benchmarks for International Real Estate Investment', *Journal of Real Estate Portfolio Management*, vol. 14 no.2, pp. 93-104, viewed 25 September 2008, EBSCOhost database.

Linneman, P 1980, 'Some Empirical Results on the Nature of Hedonic Price Functions for the Urban Housing Market', *Journal of Urban Economics*, pp. 47-68.

Maier, R, Aeberli, U, Bankert, M & Gabriel, C 2008, *Swiss Private Banking Guide 2008*, Zollikofer AG, St. Gallen, Switzerland.

Marcato, G & Key, T 2005, 'Direct Investment in Real Estate: Momentum profits and their robustness to trading costs', *Journal of Portfolio Management*, special issue 2005, pp. 55-69, viewed 25 September 2008, EBSCOhost database.

Markowitz, H 1952, 'Portfolio Selection', *Journal of Finance*, vol. 7 issue 1, March, pp. 77-91, viewed 26 July 2007, EBSCOhost database, AN 6635932.

Massheder, K & Finch, E 1998, 'Benchmarking methodologies applied to UK facilities management', *Facilities*, Vol. 16 No. 3/4, pp. 99-106.

Montezuma, J 2004, 'Housing investment in an institutional portfolio context', *Property Management*, vol. 22 no. 3, pp. 230-249, viewed 11 August 2008, <www.emeraldinsight.com/0263-7472.htm>.

Mortgage Banking 2006, ‘Global real estate markets becoming increasingly transparent’, *Mortgage Banking*, vol. 66 issue 12, September, pp. 100-2, viewed 16 October 2007, EBSCOhost database, AN 22178597.

Muller, J 2006, ‘Luring offshore money’, *Finweek Property*, p. 48, viewed 10 October 2007, EBSCOhost database, AN 21993782.

Neuman, WL 2003, *Social research methods, qualitative and quantitative approaches*, 5th edn, Allyn & Bacon, Boston, pp. 20-40, cited in USQ 2005, *MGT8401 Research Methodology 1 – Selected readings 1*, Distance and e-Learning Centre, Toowoomba, Australia, reading 1.3.

Neuman, WL 2003a, *Social research methods, qualitative and quantitative approaches*, 5th edn, Allyn & Bacon, Boston, pp. 9-12, cited in USQ 2005, *MGT8401 Research Methodology 1 – Selected readings 1*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 1.2.

Okunev, J, Wilson, P & Zurbruegg R 2002, ‘Relationships between Australian Real Estate and Stock Market Prices - a Case of Market Inefficiency’, *Journal of Forecasting*, vol. 21, October, pp. 181-92, viewed 15 October 2007, <www.interscience.wiley. com>.

Perry, C 2002, *A Structured Approach to Presenting Theses: Notes for Students and their Supervisors*, University of Southern Queensland, USQ PostGrad Toolbox, viewed 14 October 2007, < http://www.usq.edu.au/library/help/postgrad/research.htm>.

Pictet & Cie bank 1999, *BVG Index: Performance benchmark for portfolios managed on the basis of the BVG law*, Zurich, Switzerland, viewed 19 September 2008, <www.pictet.com>.

Prioni, P & Bignasca, F 2005, *Auf die Lage kommt es an*, viewed 15 October 2007, Zurich Cantonal Bank < http://intranet.prod.zkb.ch/startseite/Kunden_Maerkte/Finanzierung_Immobilien >.

Punch, KF 2006, *Developing effective research proposals*, 2d edn, SAGE Publications Ltd, London, England; Thousand Oaks, CA.

Riberio, LMM & Cobral, JAS 2006, ‘A benchmarking methodology for metal casting industry’, *Benchmarking: An International Journal*, vol. 13, pp. 23-35, viewed 19 September 2008, <www.emeraldinsight.com>.

Rom, BM & Ferguson, KW 1994, ‘Post-Modern Portfolio Theory Comes of Age’, *Journal of Investing*, vol. 3, no 3, pp. 11-17, viewed 24 August 2009, <http://www.iijournals.com/doi/abs/10.3905/joi.3.3.11>.

Rosen, S 1974, ‘Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition’, *Journal of Political Economy*, vol. 82 issue 1, February, pp. 34-55, viewed 22 September 2007, EBSCOhost database, AN 5055104.

Sarkis, J 2008, ‘Special issue on benchmarking the greening of business’, *Benchmarking: An International Journal*, vol. 15, issue 5, viewed 24 September 2008, <www.emeraldinsight.com>.

SFG - Swiss Federal Government for Statistics 2008a, *Taschenstatistik der Schweiz 2008*, SFG, Neuchâtel, Switzerland, viewed 16 May 2008, <www.bfs.admin.ch>.

SFG - Swiss Federal Government for Statistics 2008b, *Panorama 2 – Raum und Umwelt*, SFG, Neuchâtel, Switzerland, viewed 16 May 2008, <www.bfs.admin.ch>.

SFG - Swiss Federal Government for Statistics 2008c, *Bau- und Leerwohnungsstatistik der Schweiz, 2006/2007*, Swiss Federal Statistical Office, Neuchâtel, Switzerland, viewed 16 May 2008, <www.bfs.admin.ch>.

SFG - Swiss Federal Government for Statistics 2008d, *Die Berufliche Vorsorge in der Schweiz: Kennzahlen der Pensionskassenstatistik 2000-2006*, SFG, Neuchâtel, Switzerland, viewed 16 May 2008, <www.bfs.admin.ch>.

Schulte, K, Rottke, N & Pitschke, C 2005, ‘Transparency in the German real estate market’, *Benchmarking: An International Journal*, vol. 23 no. 1, pp. 90-108, viewed 19 September 2008, <www.emeraldinsight.com/1463-578X.htm>.

Sharpe, WF 1964, ‘Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk’, *Journal of Finance*, vol. 19, September, pp. 425-42, viewed 9 November 2006, < http://links.jstor.org/sici?sici=0022-1082>.

Sing, TF & Ong, SE 2000, ‘Asset Allocation in a Downside Risk Framework, *Journal of Real Estate Portfolio Management*, vol. 3, no. 3, September, pp. 213-23, viewed 19 July 2007, EBSCOhost database, AN 3543273.

STA - Statistical Office of the Canton Zurich 2008a, *Statistisches Jahrbuch des Kantons Zürich*, Druckerei Ropress AG, Zurich, Switzerland, viewed 15 April 2008, <www.statistik.zh.ch>.

STA - Statistical Office of the Canton Zurich 2008b, *Kanton Zürich in Zahlen 2008*, GDZ AG, Zurich, Switzerland, viewed 15 April 2008, <www.statistik.zh.ch>.

STA - Statistical Office of the Canton Zurich 2009, *Statistisches Jahrbuch des Kantons Zürich*, Druckerei Ropress AG, Zurich, Switzerland, viewed 1 May 2009, <www.statistik.zh.ch>.

Statsoft Inc. 2009, *Basic Statistics*, viewed 31 May 2009, available at <http://www. statsoft.com/textbook/stathome.html?stbasic.html&1 >.

Studenmund, AH 2006, ‘A Regression User’s Handbook’*, Using Econometrics: a Practical Guide*, 5th edn, Person/Addison Wesley, Boston, pp. 390-418, cited in

USQ 2007d, *MGT8402 Research Methodology 2 – Selected readings*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 5.1.

SZC - Statistical Office of the Zurich City 2008, *Statistisches Jahrbuch der Stadt Zürich*, NZZ Fretz AG, Zurich, Switzerland, viewed 19 September 2008, <www.stadt-zuerich.ch>.

Stoy, C & Schalcher, HR 2005, *Benchmarks und Einflussfaktoren der Baunutzungs-kosten*, Hochschulverlag AG an der ETH Zürich, Switzerland.

Summers, J & Smith, B 2006, *Communication Skills Handbook*, 2d edn, John Wiley & Sons Ltd, Australia.

Topintzi, E, Chin, H & Hobbs, P 2008, 'Moving towards a global real estate index', *Journal of Property Investment & Finance*, vol. 26 no. 4, pp. 286-303, viewed 19 September 2008, <www.emeraldinsight.com/1463-578X.htm>.

Trochim, WMK 2006, *Introduction to Validity*, Research Methods Knowledge Database, viewed 29 May 2007, available at http://www.socialresearchmethods.net/kb /introval.htm.

USQ - University of Southern Queensland 2006, *MGT8401 Research Methodology 1 – Study book*, Distance and e-Learning Centre, USQ, Toowoomba, Australia.

Vanini, P, Snaidero, N, Schrafl, C & Kubli, U 2008, Ausbildung: Handel und Kapitalmarkt, Grundlage Indizes, Zurich Cantonal Bank, Zurich, Switzerland, viewed 31 August 2008, <www.zkb.ch>.

W&P - Wüst & Partner AG 2007, *Immo-Monitoring: Facts & Figures 2007*, Zurich, Switzerland, viewed 15 October 2007, <http://www.wuestundpartner.com/online_services/immobilienindizes/index.phtml>.

Watson, GH 1993, *Strategic Benchmarking: How to Rate your Company's Performance against the World's Best*, Wiley, New York, NY.

Watson-Wyatt 2008, *KGAST Anlagenstiftungen: Qualität und Sicherheit für Ihre Vorsorgegelder. Performance-Vergleich (2. Säule)*, Zurich, Switzerland, viewed 19 September 2008, <www.watsonwyatt.ch>.

Weber, SC, Hasenmaile, F, Rieder, T & Sydow, N 2008, *Swiss Issues Real Estate: Real Estate Market 2008, Regions*, Credit Suisse Economic Research, Zurich, Switzerland, viewed 31 August 2008, <https://entry4.credit-suisse.ch/csfs/research/ p/d/en/publikationen/pub_immobilienstudi>.

Yin, RK 1994, *Case study research - design and methods, applied social research methods series*, vol. 5, 2nd edn, Sage, Newbury Park, London.

Zairi, M & Whymark, J 2000, 'The transfer of best practices: how build a culture of benchmarking and continuous learning – part 2’, *Benchmarking: An International Journal*, vol. 7 no. 2, pp. 146-167, viewed 19 September 2008, <www.emerald-library.com>.

Zikmund, WG 2003, ‘Bivariate analysis: measures of association, Appendices 23A’, *Business research methods*, 7th edn, Thomson, South-Western, USA, pp. 549-72, cited in USQ 2007, *MGT8402 Research Methodology 2 – Selected readings*, Distance and e-Learning Centre, USQ, Toowoomba, Australia, reading 4.1.

Zimmermann, H 2003, *Finance Compact: Fit for Finance*, Verlag Neue Zürcher Zeitung, Zürich, Switzerland.

ZKB - Zurich Cantonal Bank 2004, *Preise, Mieten und Renditen*, Zurich Cantonal Bank, Zurich, Switzerland, viewed 31 August 2007, <www.zkb.ch>.

ZKB - Zurich Cantonal Bank 2008, *Immobilienspezial: Regionale Preisentwicklung von Wohneigentum*, Zurich Cantonal Bank, Zurich, Switzerland, viewed 3 September 2008, <www.zkb.ch>.

8 APPENDICES

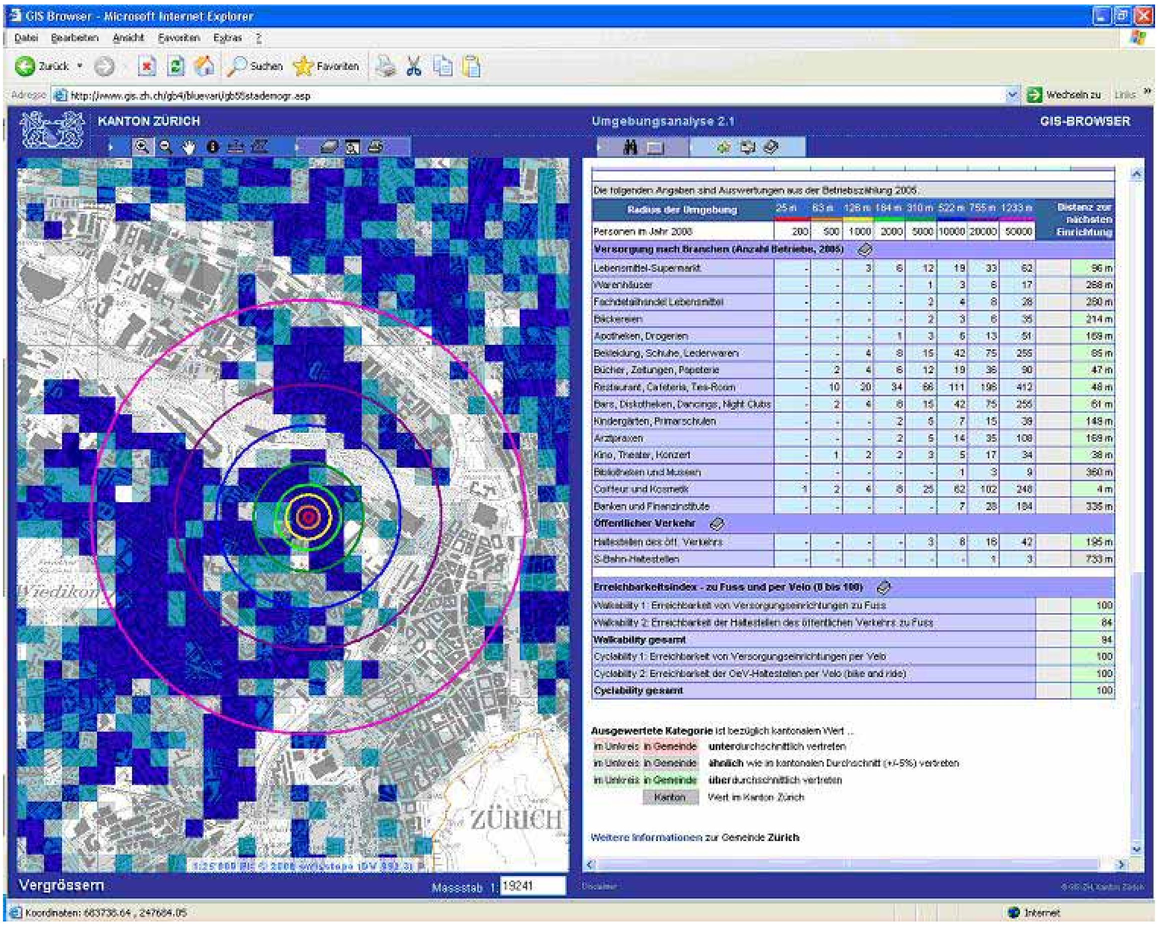
8.1 Real Estate Physical Characteristics (Variables)

Nr.	Variable	Description	Unit	Range	Transf.
DEPENDENT Variable					
-	EuclideanDist	Multivariate distance between the PFs REP_B and a specific PF REP_S	-	[0,∞]	none
INDEPENDENT Variables					
-> Structural Variables					
1	Usage	Utilisation of the RE = SFD / CO	percent	[0,1]	linear normalisation (Linear Norm)
2	Age	Age of the RE (2007-construction year)	years	[0,∞]	Linear Norm
3	NrRooms	Number of rooms per RE	rooms	[0,∞]	Linear Norm
4	Surface_m2	Usable surface per RE	square metres	[0,∞]	Linear Norm
5	Volume_m3	Volume per GVZ-Object	cubic metres	[0,∞]	Linear Norm
-> Macro Variables					
6	GeoLocationX_m	Location defined with the coordinate X	metres	[0,∞]	Linear Norm
7	GeoLocationY_m	Location defined with the coordinate Y	metres	[0,∞]	Linear Norm
8	DistanceToCentre_m	Shortest distance to centre (Zurich or Winterthur)	metres	[0,∞]	Linear Norm
9	TaxationLevel_Percent	Tax level of municipality	percent	[0,1]	Linear Norm
-> Micro Variables					
10	Lake_ha	Lake surface view	hectare	[0,∞]	Linear Norm
11	View_ha	Surface view from RE	hectare	[0,∞]	Linear Norm
12	Slope_Percent	Ground inclination	percentage	[0,1]	Linear Norm
13	SunJuly_KJ_m2	Mean of the sunshine in July 07	Kilo Joule/ square metres	[0,∞]	Linear Norm
14	Dist_Bus_m	Distance to nearest bus station	metres	[0,∞]	Linear Norm
15	Dist_Railway_m	Distance to nearest railway station	metres	[0,∞]	Linear Norm
16	Dist_School_m	Distance to nearest school	metres	[0,∞]	Linear Norm
17	Walk_Index	Walk index distance to nearest points of interest	-	[0,100]	Linear Norm

(Source: Developed for this study from available field data)

8.2 Variable "Walk_Index"

The Walk_Index is public available on the www.gis.zh.ch. The chart depicts in a graphical and numerical way the walkability from a definable centre point.



(Source: http://www.gis.zh.ch/gb4/bluevari/gb55stademogr.asp)

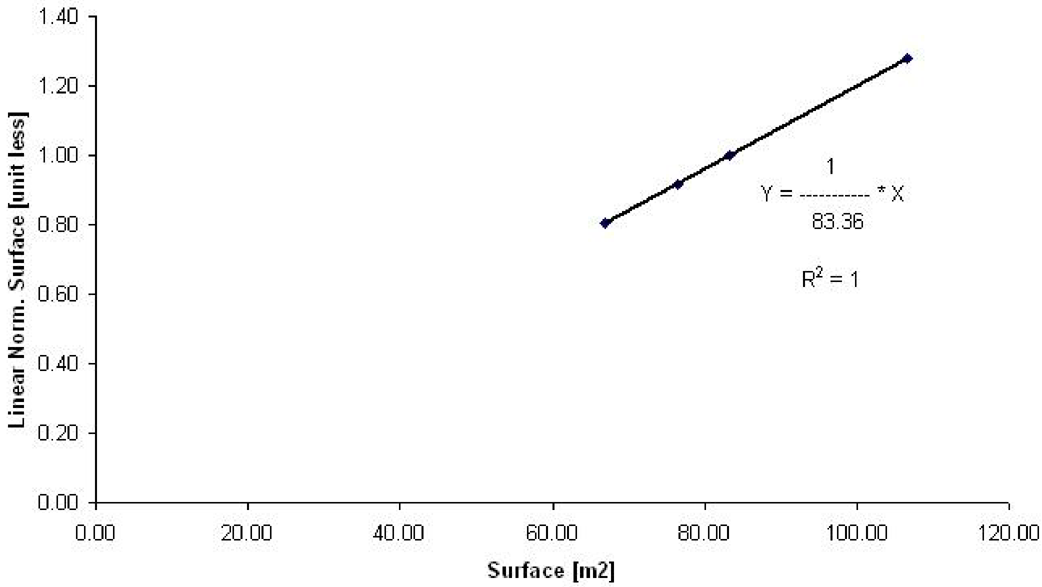
<u>Translations German -> English</u>

- Umgebungsanalyse: Surrounding area analysis.
- Erreichbarkeitsindex – zu Fuss und per Velo (1 bis 100): Reachability by foot or by bicycle (value from 0 to 100).
- Walkability gesamt: Total value for walkability.
- Cyclability gesamt: Total value for cyclability.

8.3 Linear Normalisation of the Variable "Surface"

This example of linear normalisation is calculated for the variable "Surface" and for only three PFs to show the principle used.

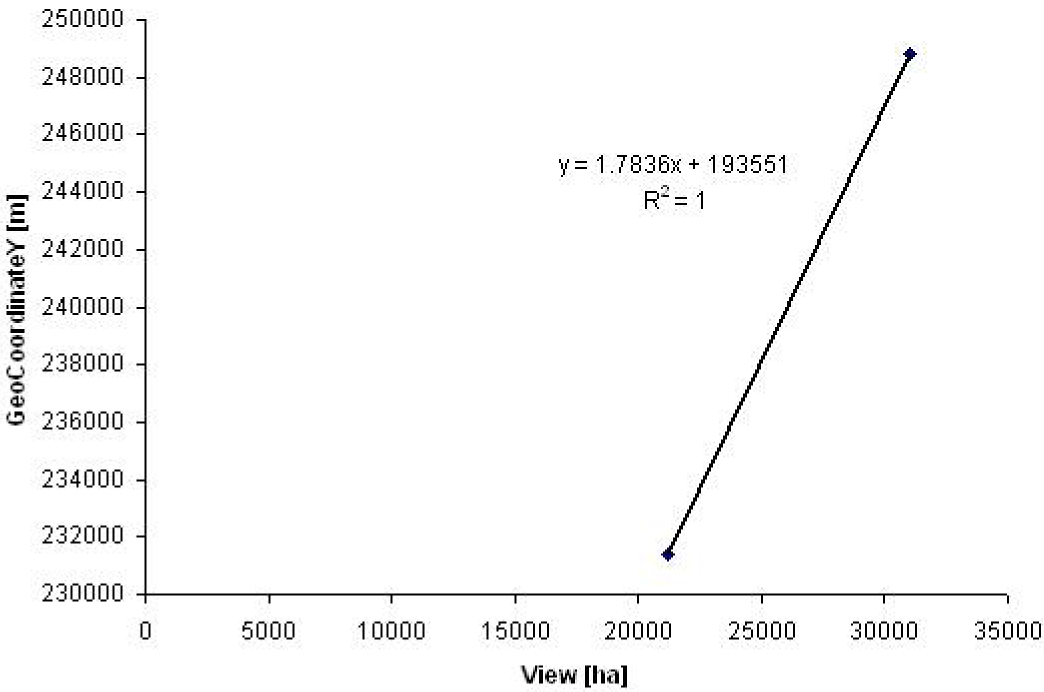
Pension Fund	PF1	PF2	PF3	Benchmark = Mean between all PFs
Mean Surface [m^2]	67.00	76.41	106.67	83.36
Linear Normalised Surface (unit less)	0.80 (67.00/83.36)	0.92 (76.41/83.36)	1.28 106.67/83.36)	1.00



(Source: Developed for this research)

8.4 Imputation Missing Values of the Variable "View"

The linear interpolation is used for the imputation of the missing value of the Variable "View."



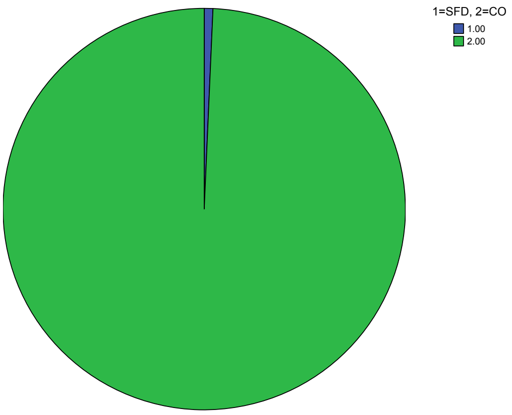
(Source: Developed from field data using SPSS)

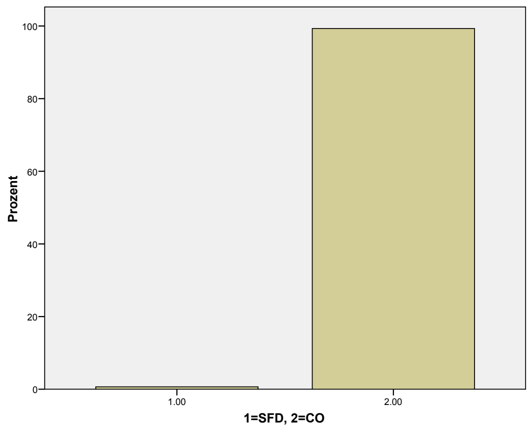
8.5 Descriptive Analysis of Variables in *REs-ZH-Dataset*, N=15,836

Following statistics are developed from field data using SPSS.

8.5.1 Analysis of the variable "Usage"- Categorical

a) Descriptive Statistic





b) Explorative Statistic - Frequency Summary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 = SFD	110	.7	.7	.7
	2 = CO	15726	99.3	99.3	100.0
	Total	15836	100.0	100.0	

c) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Usage	15836	100.0%	0	.0%	15836	100.0%

8.5.2 Analysis of the variable "Age"

a) Descriptive Statistic

			Statistic	Std. Error
Age	Mean		35.5318	.15151
	95% Confidence Interval for Mean	Lower Bound	35.2348	
		Upper Bound	35.8287	
	5% Trimmed Mean		34.7434	
	Median		35.0000	
	Variance		363.534	
	Std. Deviation		19.06656	
	Minimum		.00	
	Maximum		88.00	
	Range		88.00	
	Interquartile Range		24.00	
	Skewness		.456	.019
	Kurtosis		.252	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Age	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Age	Highest	1	224	88.00
		2	225	88.00
		3	226	88.00
		4	242	88.00
		5	243	88.00
	Lowest	1	7924	.00
		2	7923	.00
		3	7922	.00
		4	7921	.00
		5	7920	.00

8.5.3 Analysis of the variable "NrRooms"

a) Descriptive Statistic

			Statistic	Std. Error
NrRooms	Mean		3.1714	.00848
	95% Confidence Interval for Mean	Lower Bound	3.1548	
		Upper Bound	3.1880	
	5% Trimmed Mean		3.1784	
	Median		3.0000	
	Variance		1.138	
	Std. Deviation		1.06661	
	Minimum		1.00	
	Maximum		8.00	
	Range		7.00	
	Interquartile Range		1.00	
	Skewness		-.124	.019
	Kurtosis		-.053	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
NrRooms	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
NrRooms	Highest	1	7866	8.00
		2	7873	8.00
		3	2555	7.00
		4	2570	7.00
		5	2571	7.00
	Lowest	1	15785	1.00
		2	15784	1.00
		3	15768	1.00
		4	15767	1.00
		5	15766	1.00

8.5.4 Analysis of the variable "Surface"

a) Descriptive Statistic

			Statistic	Std. Error
Surface	Mean		78.5902	.21080
	95% Confidence Interval for Mean	Lower Bound	78.1770	
		Upper Bound	79.0034	
	5% Trimmed Mean		77.9276	
	Median		76.0000	
	Variance		703.679	
	Std. Deviation		26.52694	
	Minimum		12.00	
	Maximum		360.00	
	Range		348.00	
	Interquartile Range		34.00	
	Skewness		.637	.019
	Kurtosis		2.412	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Surface	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Surface	Highest	1	7866	360.00
		2	2555	300.00
		3	5526	245.00
		4	2679	242.00
		5	10102	241.00
	Lowest	1	11986	12.00
		2	13431	14.00
		3	6322	14.00
		4	11733	16.00
		5	11732	16.00

8.5.5 Analysis of the variable "Volume"

a) Descriptive Statistic

			Statistic	Std. Error
Volume	Mean		9460.2390	86.936
	95% Confidence Interval for Mean	Lower Bound	9289.8342	
		Upper Bound	9630.6439	
	5% Trimmed Mean		7702.4142	
	Median		6279.0000	
	Variance		119686938.49	
	Std. Deviation		10940.1525	
	Minimum		402.00	
	Maximum		82430.0	
	Range		82028.00	
	Interquartile Range		7247.00	
	Skewness		3.622	.019
	Kurtosis		16.266	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Volume	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Volume	Highest	1	8533	82430.0
		2	8534	82430.0
		3	8535	82430.0
		4	8536	82430.0
		5	8537	82430.0
	Lowest	1	2687	402.00
		2	8444	410.00
		3	8445	464.00
		4	4193	466.00
		5	4192	466.00

8.5.6 Analysis of the variable “GeoLocationX_m”

a) Descriptive Statistic

			Statistic	Std. Error
GeoLocationX_m	Mean		686947.2809	66.953
	95% Confidence Interval for Mean	Lower Bound	686816.0450	
		Upper Bound	687078.5168	
	5% Trimmed Mean		686759.6702	
	Median		684387.0000	
	Variance		70988548.899	
	Std. Deviation		8425.47025	
	Minimum		670439.00	
	Maximum		712095.00	
	Range		41656.00	
	Interquartile Range		13097.00	
	Skewness		.364	.019
	Kurtosis		-.745	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
GeoLocationX_m	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
GeoLocationX_m	Highest	1	11411	712095.00
		2	11412	712095.00
		3	11413	712095.00
		4	11414	712095.00
		5	11415	712095.00
	Lowest	1	4187	670439.00
		2	4186	670439.00
		3	4189	670461.00
		4	4188	670461.00
		5	4116	670561.00

8.5.7 Analysis of the variable "GeoLocationY_m"

a) Descriptive Statistic

			Statistic	Std. Error
GeoLocationY_m	Mean		249879.7676	67.903
	95% Confidence Interval for Mean	Lower Bound	249746.6692	
		Upper Bound	250012.8659	
	5% Trimmed Mean		250017.5082	
	Median		249764.0000	
	Variance		73017748.805	
	Std. Deviation		8545.04235	
	Minimum		229037.00	
	Maximum		283002.00	
	Range		53965.00	
	Interquartile Range		8514.00	
	Skewness		-.093	.019
	Kurtosis		-.143	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
GeoLocationY_m	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
GeoLocationY_m	Highest	1	5700	283002.00
		2	5701	283002.00
		3	5696	282819.00
		4	5697	282819.00
		5	5698	282819.00
	Lowest	1	7783	229037.00
		2	7782	229037.00
		3	7781	229037.00
		4	7780	229037.00
		5	7779	229037.00

8.5.8 Analysis of the variable "DistanceToCentre_m"

a) Descriptive Statistic

			Statistic	Std. Error
DistanceToCentre_m	Mean		7969.1658	45.024
	95% Confidence Interval for Mean	Lower Bound	7880.9129	
		Upper Bound	8057.4188	
	5% Trimmed Mean		7556.5342	
	Median		6682.7400	
	Variance		32102706.666	
	Std. Deviation		5665.92505	
	Minimum		379.16	
	Maximum		29848.58	
	Range		29469.42	
	Interquartile Range		7606.83	
	Skewness		1.089	.019
	Kurtosis		.698	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
DistanceToCentre_m	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
DistanceToCentre_m	Highest	1	11419	29848.58
		2	11420	29848.58
		3	11421	29848.58
		4	11422	29848.58
		5	11423	29848.58
	Lowest	1	15812	379.16
		2	15811	379.16
		3	15810	379.16
		4	6380	514.73
		5	6379	514.73

8.5.9 Analysis of the variable "TaxationLevel"

a) Descriptive Statistic

			Statistic	Std. Error
TaxationLevel	Mean		123.4610	.10875
	95% Confidence Interval for Mean	Lower Bound	123.2478	
		Upper Bound	123.6741	
	5% Trimmed Mean		124.6213	
	Median		129.5300	
	Variance		187.297	
	Std. Deviation		13.68566	
	Minimum		79.74	
	Maximum		137.94	
	Range		58.20	
	Interquartile Range		12.38	
	Skewness		-1.284	.019
	Kurtosis		.714	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
TaxationLevel	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
TaxationLevel	Highest	1	27	137.94
		2	28	137.94
		3	29	137.94
		4	30	137.94
		5	31	137.94
	Lowest	1	10241	79.74
		2	10240	79.74
		3	10239	79.74
		4	10238	79.74
		5	10237	79.74

8.5.10 Analysis of the variable "Lake_ha"

a) Descriptive Statistic

			Statistic	Std. Error
Lake_ha	Mean		514.1845	9.0218
	95% Confidence Interval for Mean	Lower Bound	496.5007	
		Upper Bound	531.8683	
	5% Trimmed Mean		324.9109	
	Median		.0000	
	Variance		1288939.506	
	Std. Deviation		1135.31472	
	Minimum		.00	
	Maximum		6523	
	Range		6523.00	
	Interquartile Range		247.00	
	Skewness		2.663	.019
	Kurtosis		7.075	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Lake_ha	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Lake_ha	Highest	1	10650	6523
		2	10651	6523
		3	10652	6523
		4	10653	6523
		5	10654	6523
	Lowest	1	15783	.00
		2	15782	.00
		3	15781	.00
		4	15780	.00
		5	15779	.00

8.5.11 Analysis of the variable "View_ha"

a) Descriptive Statistic

			Statistic	Std. Error
View_ha	Mean		14945.6758	66.637
	95% Confidence Interval for Mean	Lower Bound	14815.0592	
		Upper Bound	15076.2925	
	5% Trimmed Mean		14643.6713	
	Median		14864.0000	
	Variance		70320168.649	
	Std. Deviation		8385.71217	
	Minimum		112.00	
	Maximum		63996.00	
	Range		63884.00	
	Interquartile Range		10661.00	
	Skewness		.426	.019
	Kurtosis		.356	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
View_ha	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
View_ha	Highest	1	7862	63996.00
		2	7863	63996.00
		3	7864	63996.00
		4	7865	63996.00
		5	2554	58958.00
	Lowest	1	4191	112.00
		2	4190	112.00
		3	4213	148.00
		4	4212	148.00
		5	4211	148.00

8.5.12 Analysis of the variable "Slope"

a) Descriptive Statistic

			Statistic	Std. Error
Slope	Mean		3.0454	.02283
	95% Confidence Interval for Mean	Lower Bound	3.0006	
		Upper Bound	3.0901	
	5% Trimmed Mean		2.7634	
	Median		2.1700	
	Variance		8.256	
	Std. Deviation		2.87324	
	Minimum		.06	
	Maximum		17.98	
	Range		17.92	
	Interquartile Range		3.59	
	Skewness		1.420	.019
	Kurtosis		1.876	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Slope	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Slope	Highest	1	7849	17.98
		2	13467	15.62
		3	13468	15.62
		4	13469	15.62
		5	13470	15.62
	Lowest	1	8008	.06
		2	8007	.06
		3	8006	.06
		4	8005	.06
		5	8004	.06

8.5.13 Analysis of the variable “SunJuly”

a) Descriptive Statistic

			Statistic	Std. Error
SunJuly	Mean		5560.0767	1.11527
	95% Confidence Interval for Mean	Lower Bound	5557.8906	
		Upper Bound	5562.2628	
	5% Trimmed Mean		5550.2510	
	Median		5523.5700	
	Variance		19697.212	
	Std. Deviation		140.34676	
	Minimum		5132.02	
	Maximum		6482.20	
	Range		1350.18	
	Interquartile Range		118.28	
	Skewness		1.505	.019
	Kurtosis		3.659	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SunJuly	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
SunJuly	Highest	1	517	6482.20
		2	513	6393.13
		3	514	6393.13
		4	317	6348.77
		5	318	6348.77
	Lowest	1	13815	5132.02
		2	13814	5132.02
		3	13813	5132.02
		4	13812	5132.02
		5	13811	5132.02

8.5.14 Analysis of the variable "Dist_Bus"

a) Descriptive Statistic

			Statistic	Std. Error
Dist_Bus	Mean		169.9639	1.09600
	95% Confidence Interval for Mean	Lower Bound	167.8156	
		Upper Bound	172.1121	
	5% Trimmed Mean		156.8713	
	Median		150.6500	
	Variance		19022.302	
	Std. Deviation		137.92136	
	Minimum		8.60	
	Maximum		1976.6	
	Range		1968.02	
	Interquartile Range		115.09	
	Skewness		6.202	.019
	Kurtosis		60.039	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Dist_Bus	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Dist_Bus	Highest	1	513	1976.6
		2	514	1976.6
		3	471	1886.5
		4	472	1886.5
		5	473	1886.5
	Lowest	1	8451	8.60
		2	8450	9.05
		3	2701	12.36
		4	2700	12.36
		5	2699	12.36

8.5.15 Analysis of the variable "Dist_Railway"

a) Descriptive Statistic

			Statistic	Std. Error
Dist_Railway	Mean		847.3460	4.18405
	95% Confidence Interval for Mean	Lower Bound	839.1448	
		Upper Bound	855.5472	
	5% Trimmed Mean		807.2883	
	Median		730.8900	
	Variance		277229.133	
	Std. Deviation		526.52553	
	Minimum		17.72	
	Maximum		4252.61	
	Range		4234.89	
	Interquartile Range		638.30	
	Skewness		1.192	.019
	Kurtosis		1.816	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Dist_Railway	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Dist_Railway	Highest	1	509	4252.61
		2	7886	3889.44
		3	7887	3889.44
		4	7888	3889.44
		5	7889	3889.44
	Lowest	1	477	17.72
		2	476	17.72
		3	475	17.72
		4	474	17.72
		5	473	17.72

8.5.16 Analysis of the variable "Dist_School"

a) Descriptive Statistic

			Statistic	Std. Error
Dist_School	Mean		218.7092	1.23979
	95% Confidence Interval for Mean	Lower Bound	216.2790	
		Upper Bound	221.1393	
	5% Trimmed Mean		202.5745	
	Median		179.1000	
	Variance		24341.099	
	Std. Deviation		156.01634	
	Minimum		.00	
	Maximum		1566.9	
	Range		1566.93	
	Interquartile Range		170.94	
	Skewness		1.950	.019
	Kurtosis		5.867	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Dist_School	15836	100.0%	0	.0%	15836	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
Dist_School	Highest	1	517	1566.9
		2	513	1538.0
		3	514	1538.0
		4	82	1120.8
		5	83	1120.8
	Lowest	1	8240	.00
		2	8239	.00
		3	8238	.00
		4	8237	.00
		5	8236	.00

8.5.17 Analysis of the variable "Walk_Index"

a) Descriptive Statistic

			Statistic	Std. Error
Walk_Index	Mean		77.5980	.14167
	95% Confidence Interval for Mean	Lower Bound	77.3203	
		Upper Bound	77.8757	
	5% Trimmed Mean		78.7883	
	Median		81.1700	
	Variance		317.855	
	Std. Deviation		17.82850	
	Minimum		3.11	
	Maximum		100.0	
	Range		96.89	
	Interquartile Range		23.50	
	Skewness		-.915	.019
	Kurtosis		.271	.039

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Walk_Index	15836	100.0%	0	.0%	15836	100.0%

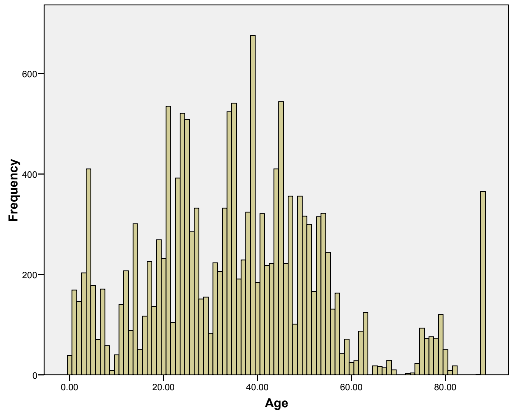
c) Explorative Statistic - Extreme Values

			Case Number	Value
Walk_Index	Highest	1	359	100.0
		2	360	100.0
		3	361	100.0
		4	362	100.0
		5	363	100.0
	Lowest	1	514	3.11
		2	513	3.11
		3	323	5.98
		4	322	5.98
		5	321	5.98

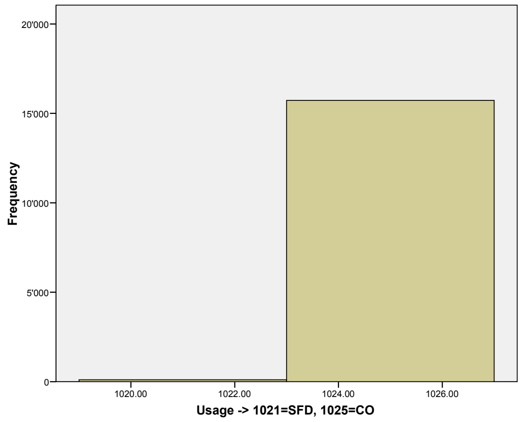
8.6 Histograms of Variables in *REs-ZH-Dataset*, N=15,836

Following histograms are developed from field data using SPSS.

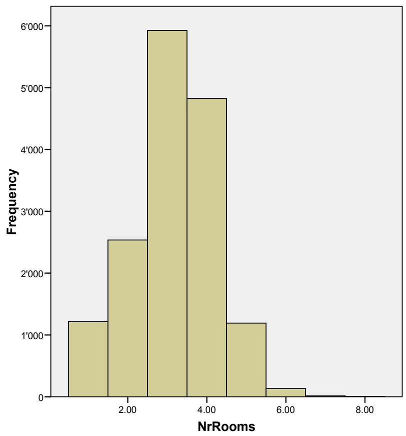
1- Age



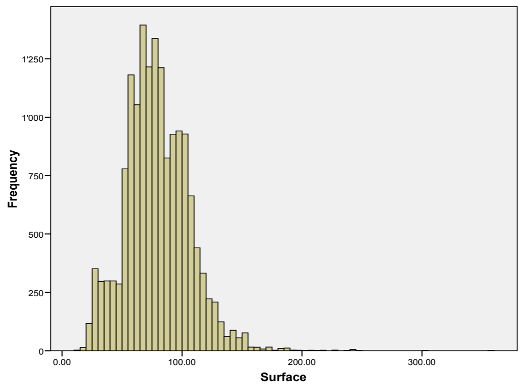
2- Usage



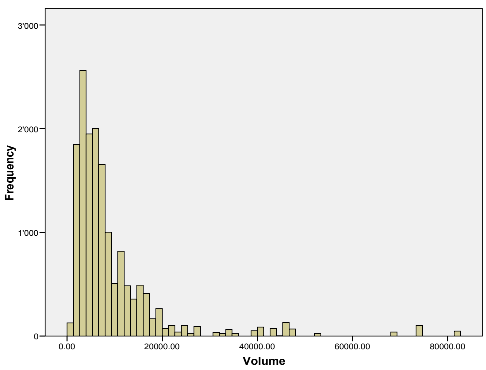
3- NrRooms



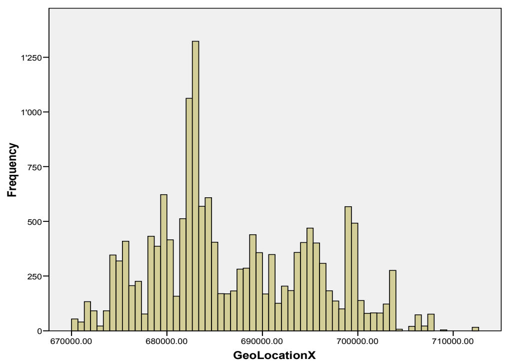
4- Surface



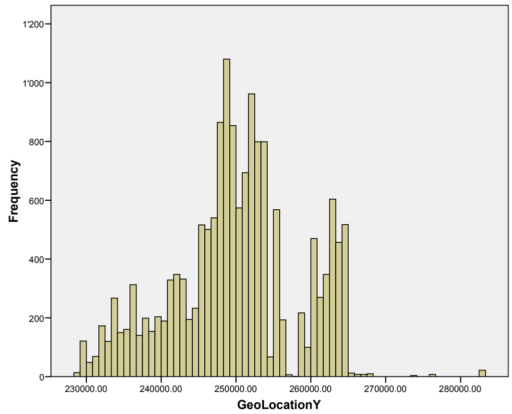
5- Volume



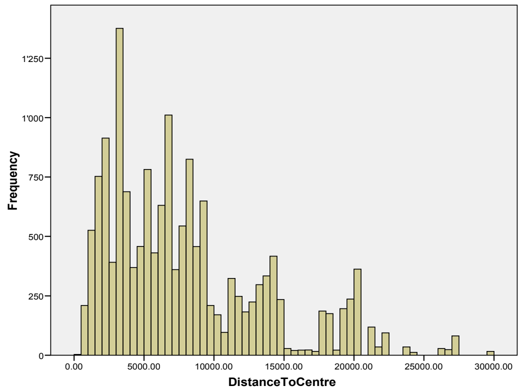
6- GeoLocationX



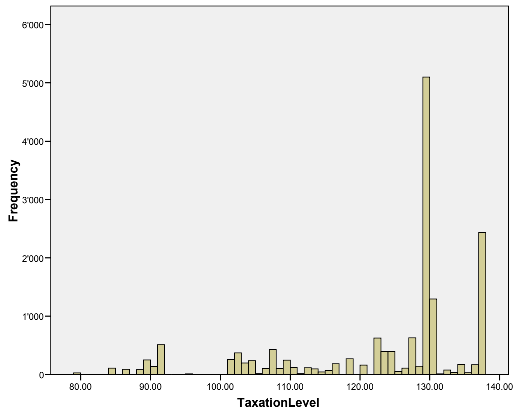
7- GeoLocationY



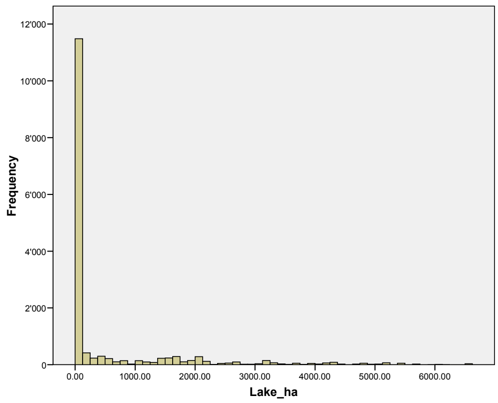
8- DistanceToCentre



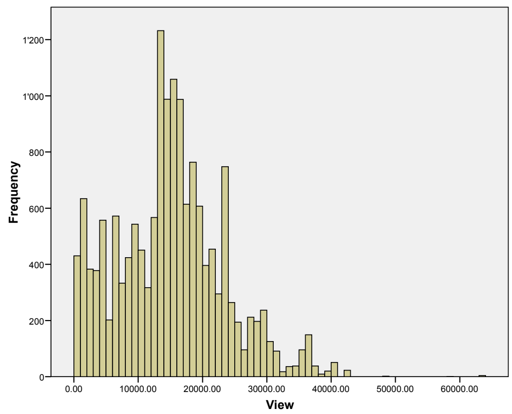
9- TaxationLevel



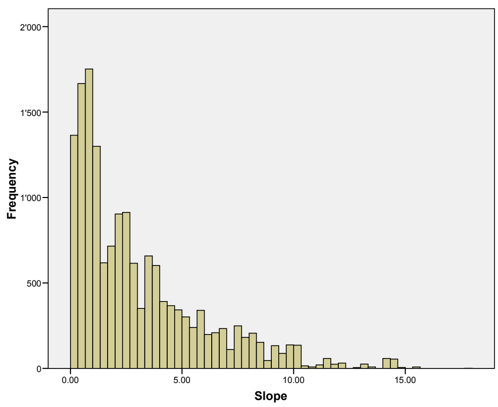
10- Lake



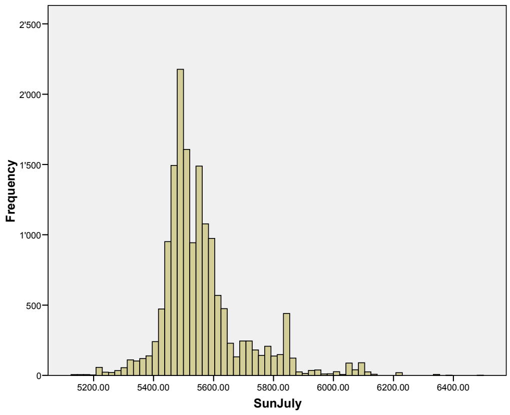
11- View



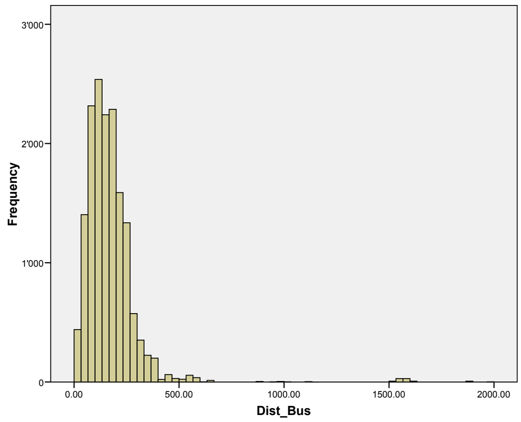
12- Slope



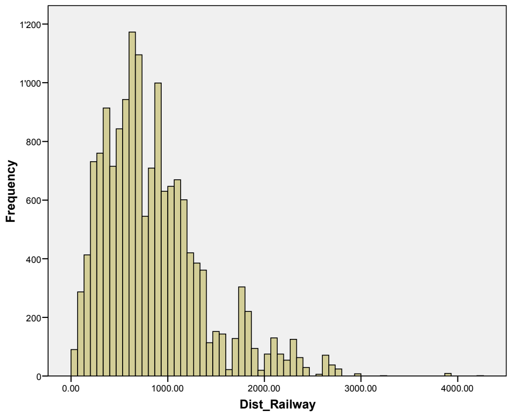
13- SunJuly



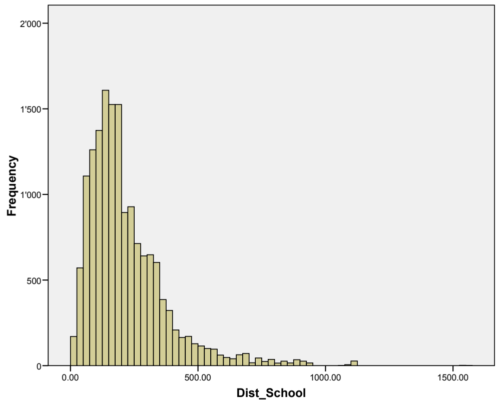
14- Dist_Bus



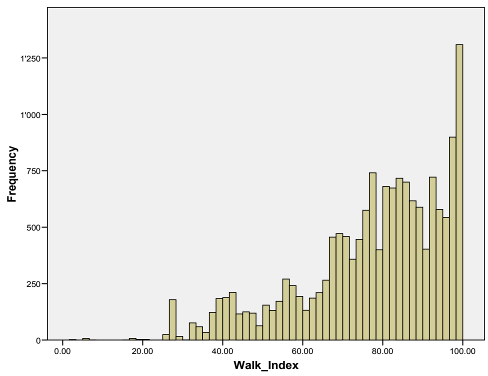
15- Dist_Railway



16- Dist_School



17- Walk_Index



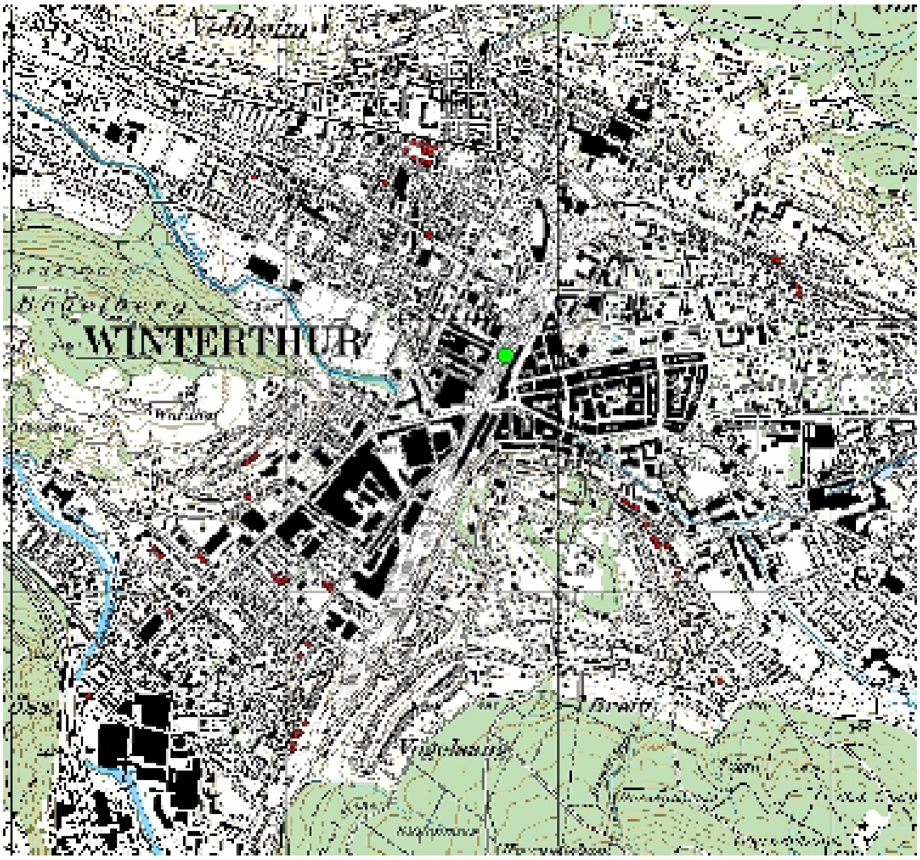
8.7 Paradeplatz as Geographical Centre Point of Zurich

Paradeplatz is considered the centre point (green dot) for Zurich with the geo-coordinates X equal to 683,101 and Y equal to 247,124 metres.

(Source: Developed from field data using GIS)

8.8 Railway Station as Geographical Centre Point of Winterthur

The railway station is considered the centre point (green dot) for Winterthur with the geo-coordinate X equal to 696,806 and Y equal to 261,869 metres.



(Source: Developed from field data using GIS)

8.9 Descriptive Statistic "Surface" with "NrRooms" as Factor

	NrRooms			Statistic	Std. Error
Surface_m2	1.00	Mean		34.3171	.34993
		95% Confidence Interval for Mean	Lower Bound	33.6306	
			Upper Bound	35.0037	
		5% Trimmed Mean		32.8603	
		Median		33.0000	
		Variance		148.655	
		Std. Deviation		12.19243	
		Minimum		12.00	
		Maximum		104.00	
		Range		92.00	
		Interquartile Range		9.00	
		Skewness		2.809	.070
		Kurtosis		11.321	.140
	2.00	Mean		57.6852	.23051
		95% Confidence Interval for Mean	Lower Bound	57.2332	
			Upper Bound	58.1372	
		5% Trimmed Mean		57.2954	
		Median		56.0000	
		Variance		134.698	
		Std. Deviation		11.60595	
		Minimum		20.00	
		Maximum		180.00	
		Range		160.00	
		Interquartile Range		15.00	
		Skewness		1.110	.049
		Kurtosis		7.015	.097
	3.00	Mean		74.0848	.18239
		95% Confidence Interval for Mean	Lower Bound	73.7272	
			Upper Bound	74.4423	
		5% Trimmed Mean		73.4336	
		Median		74.0000	
		Variance		197.028	
		Std. Deviation		14.03666	
		Minimum		25.00	
		Maximum		242.00	

	Range		217.00	
	Interquartile Range		18.00	
	Skewness		1.120	.032
	Kurtosis		5.482	.064
4.00	Mean		94.4681	.24408
	95% Confidence Interval for Mean	Lower Bound	93.9896	
		Upper Bound	94.9466	
	5% Trimmed Mean		94.0620	
	Median		95.0000	
	Variance		287.261	
	Std. Deviation		16.94877	
	Minimum		40.00	
	Maximum		198.00	
	Range		158.00	
	Interquartile Range		22.00	
	Skewness		.446	.035
	Kurtosis		1.392	.071
5.00	Mean		117.2729	.57977
	95% Confidence Interval for Mean	Lower Bound	116.1354	
		Upper Bound	118.4104	
	5% Trimmed Mean		116.1692	
	Median		112.0000	
	Variance		400.328	
	Std. Deviation		20.00820	
	Minimum		56.00	
	Maximum		245.00	
	Range		189.00	
	Interquartile Range		25.00	
	Skewness		1.267	.071
	Kurtosis		3.828	.142
6.00	Mean		144.2256	2.60817
	95% Confidence Interval for Mean	Lower Bound	139.0664	
		Upper Bound	149.3848	
	5% Trimmed Mean		141.6566	
	Median		139.0000	
	Variance		904.737	
	Std. Deviation		30.07884	
	Minimum		85.00	
	Maximum		241.00	
	Range		156.00	

	Interquartile Range		31.00	
	Skewness		1.275	.210
	Kurtosis		2.160	.417
7.00	Mean		179.8750	10.37098
	95% Confidence Interval for Mean	Lower Bound	157.7698	
		Upper Bound	201.9802	
	5% Trimmed Mean		178.0278	
	Median		170.0000	
	Variance		1720.917	
	Std. Deviation		41.48393	
	Minimum		93.00	
	Maximum		300.00	
	Range		207.00	
	Interquartile Range		18.00	
	Skewness		1.124	.564
	Kurtosis		5.352	1.091
8.00	Mean		300.0000	60.00000
	95% Confidence Interval for Mean	Lower Bound	-462.3723	
		Upper Bound	1062.3723	
	5% Trimmed Mean		.	
	Median		300.0000	
	Variance		7200.000	
	Std. Deviation		84.85281	
	Minimum		240.00	
	Maximum		360.00	
	Range		120.00	
	Interquartile Range		.	
	Skewness		.	.
	Kurtosis		.	.

(Source: Developed from field data using SPSS)

8.10 Crisscross Table for PFs in *REs-ZH-Dataset*, N=74

The structure of the crisscross table is depicted in the following table. The values of the mean and the standard deviation have been calculated for each PF and for each REP variable.

	REs	Age		NrRooms		…	
PF_CodeNr	Number	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
1.00	10	50.00	.00	3.20	.42		
2.00	213	32.56	8.48	3.55	1.07		
3.00	3	88.00	.00	3.67	.58		
4.00	12	38.42	4.48	3.75	1.36		
5.00	3	32.00	.00	4.00	.00		
6.00	9	88.00	.00	3.67	1.22		
7.00	66	36.05	2.38	3.61	.80		
8.00	11	28.36	17.15	3.09	.94		
9.00	194	32.75	21.02	3.33	1.16		
10.00	59	54.03	17.54	3.03	.83		
11.00	5	48.00	.00	3.00	1.87		
12.00	1836	35.61	14.66	3.42	.79		
13.00	16	60.63	19.55	3.19	1.22		
14.00	63	40.51	15.90	2.84	1.17		
…	…	…	…	…	…		
72.00	103	42.97	21.54	2.83	1.27		
73.00	3	56.00	.00	3.67	1.15		
74.00	24	34.00	.00	3.50	.51		

(Source: Developed from field data using SPSS)

8.11 Outliers by Dependent Variable "EuclideanDist", N=74

Following statistics are developed from field data using SPSS.

a) Descriptive Statistic

			Statistic	Std. Error
EuclideanDist	Mean		2.9236	.42868
	95% Confidence Interval for Mean	Lower Bound	2.0693	
		Upper Bound	3.7780	
	5% Trimmed Mean		2.3962	
	Median		2.0611	
	Variance		13.599	
	Std. Deviation		3.68765	
	Minimum		.77	
	Maximum		31.37	
	Range		30.59	
	Interquartile Range		1.03	
	Skewness		6.612	.279
	Kurtosis		49.877	.552

b) Explorative Statistic - Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
EuclideanDist	74	100.0%	0	.0%	74	100.0%

c) Explorative Statistic - Extreme Values

			Case Number	Value
EuclideanDist	Highest	1	55	31.37
		2	38	10.73
		3	62	6.60
		4	1	6.23
		5	19	6.02
	Lowest	1	10	.77
		2	69	.91
		3	18	1.14
		4	68	1.26
		5	67	1.29

8.12 Test of Normality for Variables in the *PFs-ZH-Dataset*, N=72

Following statistics are developed from field data using SPSS.

a) Kolmogorov-Smirnov and Shapiro-Wilk

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
EuclideanDist	.204	72	.000	.806	72	.000
Usage	.394	72	.000	.432	72	.000
Age	.143	72	.001	.939	72	.002
NrRooms	.161	72	.000	.886	72	.000
Surface_m2	.129	72	.005	.920	72	.000
Volume_m3	.145	72	.001	.799	72	.000
GeoLocationX_m	.090	72	.200(*)	.976	72	.175
GeoLocationY_m	.097	72	.089	.962	72	.029
DistanceToCentre_m	.150	72	.000	.939	72	.002
TaxationLevel_Percent	.154	72	.000	.874	72	.000
Lake_ha	.257	72	.000	.693	72	.000
View_ha	.095	72	.178	.963	72	.032
Slope_Percent	.182	72	.000	.815	72	.000
SunJuli_KJ_m2	.170	72	.000	.862	72	.000
Dist_Bus_m	.198	72	.000	.623	72	.000
Dist_Railway_m	.203	72	.000	.768	72	.000
Dist_School_m	.175	72	.000	.867	72	.000
Walk_Index	.174	72	.000	.858	72	.000

* This is a lower bound of the true significance.
a Lilliefors Significance Correction

8.13 Descriptive Analysis of Variables in *PFs-ZH-Dataset*, N=71

Following statistics are developed from field data using SPSS.

			Statistic	Std. Error
EuclideanDist	Mean		2.4235	.14487
	95% Confidence Interval for Mean	Lower Bound	2.1345	
		Upper Bound	2.7124	
	5% Trimmed Mean		2.2942	
	Median		2.0423	
	Variance		1.490	
	Std. Deviation		1.22071	
	Minimum		.77	
	Maximum		6.60	
	Range		5.82	
	Interquartile Range		1.02	
	Skewness		1.796	.285
	Kurtosis		3.164	.563
LN_EuclideanDist	Mean		.7866	.05103
	95% Confidence Interval for Mean	Lower Bound	.6848	
		Upper Bound	.8884	
	5% Trimmed Mean		.7743	
	Median		.7141	
	Variance		.185	
	Std. Deviation		.42998	
	Minimum		-.26	
	Maximum		1.89	
	Range		2.14	
	Interquartile Range		.48	
	Skewness		.559	.285
	Kurtosis		.629	.563
Usage	Mean		.9925	.31720
	95% Confidence Interval for Mean	Lower Bound	.3599	
		Upper Bound	1.6251	
	5% Trimmed Mean		.4982	
	Median		.0000	
	Variance		7.144	
	Std. Deviation		2.67274	
	Minimum		.00	
	Maximum		15.38	
	Range		15.38	
	Interquartile Range		.12	
	Skewness		3.571	.285
	Kurtosis		13.989	.563

Age	Mean		42.2882	2.47325
	95% Confidence Interval for Mean	Lower Bound	37.3555	
		Upper Bound	47.2209	
	5% Trimmed Mean		41.9317	
	Median		39.0741	
	Variance		434.305	
	Std. Deviation		20.83998	
	Minimum		1.72	
	Maximum		88.00	
	Range		86.28	
	Interquartile Range		19.43	
	Skewness		.539	.285
	Kurtosis		.288	.563
NrRooms	Mean		3.3614	.08787
	95% Confidence Interval for Mean	Lower Bound	3.1861	
		Upper Bound	3.5366	
	5% Trimmed Mean		3.3184	
	Median		3.2636	
	Variance		.548	
	Std. Deviation		.74042	
	Minimum		1.38	
	Maximum		6.00	
	Range		4.62	
	Interquartile Range		.61	
	Skewness		1.247	.285
	Kurtosis		4.018	.563
Surface_m2	Mean		86.3683	2.54910
	95% Confidence Interval for Mean	Lower Bound	81.2843	
		Upper Bound	91.4523	
	5% Trimmed Mean		85.0736	
	Median		81.2452	
	Variance		461.353	
	Std. Deviation		21.47913	
	Minimum		36.63	
	Maximum		150.00	
	Range		113.37	
	Interquartile Range		23.72	
	Skewness		1.031	.285
	Kurtosis		1.460	.563
Volume_m3	Mean		5420.7801	512.60279
	95% Confidence Interval for Mean	Lower Bound	4398.4261	
		Upper Bound	6443.1341	
	5% Trimmed Mean		4942.3775	
	Median		4236.9474	
	Variance		18656074.980	

	Std. Deviation		4319.26788	
	Minimum		775.25	
	Maximum		26262.5	
	Range		25487.27	
	Interquartile Range		4353.82	
	Skewness		2.294	.285
	Kurtosis		7.796	.563
GeoLocationX_m	Mean		687531.3106	950.02260
	95% Confidence Interval for Mean	Lower Bound	685636.5503	
		Upper Bound	689426.0709	
	5% Trimmed Mean		687515.6044	
	Median		686560.0000	
	Variance		64080548.235	
	Std. Deviation		8005.03268	
	Minimum		670650.36	
	Maximum		705200.27	
	Range		34549.91	
	Interquartile Range		10085.00	
	Skewness		.191	.285
	Kurtosis		-.358	.563
GeoLocationY_m	Mean		248825.1752	1174.76631
	95% Confidence Interval for Mean	Lower Bound	246482.1776	
		Upper Bound	251168.1727	
	5% Trimmed Mean		248473.6669	
	Median		248090.0000	
	Variance		97985387.769	
	Std. Deviation		9898.75688	
	Minimum		230549.33	
	Maximum		282880.00	
	Range		52330.67	
	Interquartile Range		11496.69	
	Skewness		.650	.285
	Kurtosis		1.267	.563
DistanceToCentre_m	Mean		9929.4531	686.52114
	95% Confidence Interval for Mean	Lower Bound	8560.2299	
		Upper Bound	11298.6763	
	5% Trimmed Mean		9784.5905	
	Median		8790.8663	
	Variance		33463100.832	
	Std. Deviation		5784.72997	
	Minimum		379.16	
	Maximum		22210.03	
	Range		21830.87	
	Interquartile Range		7908.19	
	Skewness		.533	.285

	Kurtosis		-.595	.563
TaxationLevel_Percent	Mean		121.6335	1.59899
	95% Confidence Interval for Mean	Lower Bound	118.4444	
		Upper Bound	124.8226	
	5% Trimmed Mean		122.6946	
	Median		123.8200	
	Variance		181.530	
	Std. Deviation		13.47332	
	Minimum		84.17	
	Maximum		137.94	
	Range		53.77	
	Interquartile Range		13.70	
	Skewness		-1.247	.285
	Kurtosis		1.137	.563
Lake_ha	Mean		792.1803	147.01412
	95% Confidence Interval for Mean	Lower Bound	498.9698	
		Upper Bound	1085.3907	
	5% Trimmed Mean		613.9009	
	Median		278.9500	
	Variance		1534533.852	
	Std. Deviation		1238.76303	
	Minimum		.00	
	Maximum		5621	
	Range		5620.55	
	Interquartile Range		914.47	
	Skewness		2.244	.285
	Kurtosis		5.240	.563
View_ha	Mean		16068.8266	789.07750
	95% Confidence Interval for Mean	Lower Bound	14495.0612	
		Upper Bound	17642.5921	
	5% Trimmed Mean		15890.3804	
	Median		15455.8768	
	Variance		44207673.940	
	Std. Deviation		6648.88517	
	Minimum		1635.60	
	Maximum		39397.65	
	Range		37762.05	
	Interquartile Range		7385.94	
	Skewness		.515	.285
	Kurtosis		1.701	.563
Slope_Percent	Mean		3.5881	.32459
	95% Confidence Interval for Mean	Lower Bound	2.9408	
		Upper Bound	4.2355	
	5% Trimmed Mean		3.3121	

	Median		2.6900	
	Variance		7.481	
	Std. Deviation		2.73508	
	Minimum		.56	
	Maximum		14.94	
	Range		14.38	
	Interquartile Range		2.78	
	Skewness		1.852	.285
	Kurtosis		4.022	.563
SunJuli_KJ_m2	Mean		5580.8610	17.24624
	95% Confidence Interval for Mean	Lower Bound	5546.4644	
		Upper Bound	5615.2575	
	5% Trimmed Mean		5570.6466	
	Median		5550.2572	
	Variance		21117.736	
	Std. Deviation		145.31943	
	Minimum		5280.21	
	Maximum		6208.45	
	Range		928.24	
	Interquartile Range		121.90	
	Skewness		1.647	.285
	Kurtosis		4.726	.563
Dist_Bus_m	Mean		170.6932	14.41966
	95% Confidence Interval for Mean	Lower Bound	141.9341	
		Upper Bound	199.4523	
	5% Trimmed Mean		157.2656	
	Median		150.1328	
	Variance		14762.784	
	Std. Deviation		121.50220	
	Minimum		20.22	
	Maximum		983.31	
	Range		963.09	
	Interquartile Range		74.64	
	Skewness		4.492	.285
	Kurtosis		28.600	.563
Dist_Railway_m	Mean		884.1523	64.71907
	95% Confidence Interval for Mean	Lower Bound	755.0742	
		Upper Bound	1013.2304	
	5% Trimmed Mean		828.9274	
	Median		822.5458	
	Variance		297387.587	
	Std. Deviation		545.33255	
	Minimum		79.07	
	Maximum		3889.44	
	Range		3810.37	
	Interquartile Range		388.01	

	Skewness		2.819	.285
	Kurtosis		12.789	.563
Dist_School_m	Mean		243.9126	14.42667
	95% Confidence Interval for Mean	Lower Bound	215.1395	
		Upper Bound	272.6857	
	5% Trimmed Mean		234.2975	
	Median		222.3829	
	Variance		14777.136	
	Std. Deviation		121.56124	
	Minimum		29.83	
	Maximum		808.02	
	Range		778.19	
	Interquartile Range		108.43	
	Skewness		1.848	.285
	Kurtosis		5.896	.563
Walk_Index	Mean		76.6955	1.93672
	95% Confidence Interval for Mean	Lower Bound	72.8329	
		Upper Bound	80.5582	
	5% Trimmed Mean		78.1852	
	Median		80.2833	
	Variance		266.312	
	Std. Deviation		16.31905	
	Minimum		13.13	
	Maximum		100.0	
	Range		86.87	
	Interquartile Range		13.95	
	Skewness		-1.714	.285
	Kurtosis		4.088	.563

8.14 Number of REs per PFs in Canton Zurich

Following statistics are developed from field data using SPSS.

PF_CodeNr	Number REs	Percent	PF_CodeNr	Number REs	Percent
66	5022	31.79%	16	22	0.14%
12	1836	11.62%	22	20	0.13%
67	1674	10.60%	62	20	0.13%
39	1275	8.07%	21	17	0.11%
58	989	6.26%	61	17	0.11%
26	839	5.31%	13	16	0.10%
64	440	2.79%	24	16	0.10%
69	353	2.23%	32	16	0.10%
48	349	2.21%	33	16	0.10%
23	258	1.63%	70	16	0.10%
2	213	1.35%	19	15	0.09%
51	208	1.32%	44	14	0.09%
9	194	1.23%	57	13	0.08%
68	177	1.12%	4	12	0.08%
60	164	1.04%	8	11	0.07%
18	133	0.84%	41	11	0.07%
42	125	0.79%	1	10	0.06%
59	110	0.70%	27	9	0.06%
72	103	0.65%	30	9	0.06%
65	100	0.63%	63	9	0.06%
49	92	0.58%	28	8	0.05%
37	81	0.51%	36	6	0.04%
29	79	0.50%	40	6	0.04%
7	66	0.42%	43	6	0.04%
14	63	0.40%	46	6	0.04%
47	63	0.40%	11	5	0.03%
53	63	0.40%	15	5	0.03%
10	59	0.37%	35	5	0.03%
20	58	0.37%	34	4	0.03%
52	56	0.35%	50	4	0.03%
25	46	0.29%	3	3	0.02%
45	45	0.28%	5	3	0.02%
71	30	0.19%	73	3	0.02%
54	27	0.17%	56	2	0.01%
17	26	0.16%	31	1	0.01%
74	24	0.15%			
			Total	15796	100.00%

8.15 MRE - All Predictors Entered Simultaneously (ED)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics

	Mean	Std. Deviation	N
EuclideanDist	2.4235	1.22071	71
Surface_m2	86.3683	21.47913	71
Lake_ha	792.1803	1238.76303	71
Usage	.9925	2.67274	71
DistanceToCentre_m	9929.4531	5784.72997	71
Dist_Railway_m	884.1523	545.33255	71
Age	42.2882	20.83998	71
NrRooms	3.3614	.74042	71
Volume_m3	5420.7801	4319.26788	71
GeoLocationX_m	687531.31	8005.03268	71
GeoLocationY_m	248825.17	9898.75688	71
TaxationLevel_Percent	121.6335	13.47332	71
View_ha	16068.826	6648.88517	71
Slope_Percent	3.5881	2.73508	71
SunJuli_KJ_m2	5580.8610	145.31943	71
Dist_Bus_m	170.6932	121.50220	71
Dist_School_m	243.9126	121.56124	71
Walk_Index	76.6955	16.31905	71

b) Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.856(a)	.732	.646	.72603	1.847

c) ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	76.371	17	4.492	8.523	.000(a)
	Residual	27.937	53	.527		
	Total	104.309	70			

(a) Predictors: (Constant), Walk_Index, NrRooms, Lake_ha, View_ha, Volume_m3, GeoLocationX_m, Usage, Dist_School_m, DistanceToCentre_m, Dist_Railway_m, Surface_m2, TaxationLevel_Percent, Slope_Percent, Dist_Bus_m, Age, SunJuli_KJ_m2, GeoLocationY_m,
(b) Dependent Variable: EuclideanDist

d) Coefficients(a)

Model	Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tol.	VIF
(Constant)	4.880	10.463		.466	.643		
Surface_m2	.012	.008	.212	1.546	.128	.268	3.731
Lake_ha	.001	.000	.646	4.910	.000	.292	3.426
Usage	.203	.037	.444	5.452	.000	.761	1.315
DistanceToCentre_m	3.15E-005	.000	.149	1.629	.109	.600	1.666
Dist_Railway_m	.000	.000	.126	1.351	.182	.585	1.709
Age	-.001	.006	-.022	-.229	.819	.569	1.756
NrRooms	-.060	.212	-.037	-.284	.777	.305	3.279
Volume_m3	1.82E-006	.000	.006	.076	.940	.704	1.421
GeoLocationX_m	-1.02E-005	.000	-.067	-.718	.476	.579	1.727
GeoLocationY_m	4.91E-006	.000	.040	.343	.733	.374	2.672
TaxationLevel_Percent	.011	.009	.121	1.277	.207	.562	1.779
View_ha	4.11E-006	.000	.022	.259	.796	.677	1.477
Slope_Percent	.018	.045	.039	.390	.698	.495	2.019
SunJuli_KJ_m2	7.40E-005	.001	.009	.084	.934	.456	2.192
Dist_Bus_m	-7.57E-005	.001	-.008	-.075	.940	.507	1.974
Dist_School_m	5.82E-005	.001	.006	.069	.945	.712	1.405
Walk_Index	-.008	.007	-.103	-1.082	.284	.556	1.800

(a) Dependent Variable: EuclideanDist

8.16 MRE - All Predictors Entered Stepwise (ED)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics

	Mean	Std. Deviation	N
EuclideanDist	2.4235	1.22071	71
Surface_m2	86.3683	21.47913	71
Lake_ha	792.1803	1238.76303	71
Usage	.9925	2.67274	71
DistanceToCentre_m	9929.4531	5784.72997	71
Dist_Railway_m	884.1523	545.33255	71
Age	42.2882	20.83998	71
NrRooms	3.3614	.74042	71
Volume_m3	5420.7801	4319.26788	71
GeoLocationX_m	687531.31	8005.03268	71
GeoLocationY_m	248825.17	9898.75688	71
TaxationLevel_Percent	121.6335	13.47332	71
View_ha	16068.826	6648.88517	71
Slope_Percent	3.5881	2.73508	71
SunJuli_KJ_m2	5580.8610	145.31943	71
Dist_Bus_m	170.6932	121.50220	71
Dist_School_m	243.9126	121.56124	71
Walk_Index	76.6955	16.31905	71

b) Model Summary(f)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.556(a)	.309	.299	1.02220	
2	.762(b)	.581	.569	.80167	
3	.808(c)	.652	.637	.73587	
4	.830(d)	.690	.671	.70051	
5	.841(e)	.708	.685	.68487	1.786

c) ANOVA(f)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	32.212	1	32.212	30.828	.000(a)
	Residual	72.097	69	1.045		
	Total	104.309	70			
2	Regression	60.607	2	30.303	47.152	.000(b)
	Residual	43.702	68	.643		
	Total	104.309	70			
3	Regression	68.028	3	22.676	41.876	.000(c)
	Residual	36.281	67	.542		
	Total	104.309	70			

4	Regression	71.921	4	17.980	36.641	.000(d)
	Residual	32.387	66	.491		
	Total	104.309	70			
5	Regression	73.820	5	14.764	31.476	.000(e)
	Residual	30.488	65	.469		
	Total	104.309	70			

(a) Predictors: (Constant), Lake_ha
(b) Predictors: (Constant), Lake_ha, Usage
(c) Predictors: (Constant), Lake_ha, Usage, DistanceToCentre_m
(d) Predictors: (Constant), Lake_ha, Usage, DistanceToCentre_m, Surface_m2
(e) Predictors: (Constant), Lake_ha, Usage, DistanceToCentre_m, Surface_m2, Dist_Railway_m
(f) Dependent Variable: EuclideanDist

d) Coefficients(a)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	(Constant)	1.990	.144		13.789	.000		
	Lake_ha	.001	.000	.556	5.552	.000	1.000	1.000
2	(Constant)	1.741	.119		14.602	.000		
	Lake_ha	.001	.000	.572	7.279	.000	.999	1.001
	Usage	.238	.036	.522	6.647	.000	.999	1.001
3	(Constant)	1.208	.181		6.685	.000		
	Lake_ha	.001	.000	.546	7.536	.000	.990	1.010
	Usage	.227	.033	.498	6.876	.000	.991	1.009
	DistanceToCentre_m	5.68E-005	.000	.269	3.702	.000	.983	1.017
4	(Constant)	.324	.358		.906	.368		
	Lake_ha	.001	.000	.563	8.135	.000	.982	1.018
	Usage	.215	.032	.471	6.766	.000	.972	1.029
	DistanceToCentre_m	4.46E-005	.000	.211	2.929	.005	.904	1.106
	Surface_m2	.012	.004	.204	2.817	.006	.893	1.120
5	(Constant)	.149	.361		.414	.680		
	Lake_ha	.001	.000	.581	8.511	.000	.966	1.036
	Usage	.212	.031	.464	6.820	.000	.970	1.031
	DistanceToCentre_m	3.83E-005	.000	.182	2.520	.014	.866	1.155
	Surface_m2	.011	.004	.194	2.726	.008	.888	1.126
	Dist_Railway_m	.000	.000	.140	2.012	.048	.923	1.083

(a) Dependent Variable: EuclideanDist

8.17 MRE - All Predictors Entered Stepwise (ED, no Constant)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics(b)

	Mean(a)	Root Mean Square	N
EuclideanDist	2.4235	2.70967	71
Surface_m2	86.3683	88.96258	71
Lake_ha	792.1803	1463.03461	71
Usage	.9925	2.83338	71
DistanceToCentre_m	9929.4531	11471.0866	71
Dist_Railway_m	884.1523	1036.78556	71
Age	42.2882	47.07950	71
NrRooms	3.3614	3.44084	71
Volume_m3	5420.7801	6912.17550	71
GeoLocationX_m	687531.31	687577.254	71
GeoLocationY_m	248825.1752	249019.222	71
TaxationLevel_Percent	121.6335	122.36697	71
View_ha	16068.826	17372.1679	71
Slope_Percent	3.5881	4.50001	71
SunJuli_KJ_m2	5580.8610	5582.72597	71
Dist_Bus_m	170.6932	209.02401	71
Dist_School_m	243.9126	272.14400	71
Walk_Index	76.6955	78.38855	71

(a) The observed mean is printed
(b) Coefficients have been calculated through the origin.

b) Model Summary(g,h)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.901(b)	.812	.809	1.18401	
2	.940(c)	.884	.881	.93558	
3	.963(d)	.928	.925	.74141	
4	.968(e)	.937	.933	.69958	
5	.970(f)	.941	.937	.68056	1.802

(a) For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared to R Square for models which include an intercept.
(b) Predictors: Surface_m2
(c) Predictors: Surface_m2, Lake_ha
(d) Predictors: Surface_m2, Lake_ha, Usage
(e) Predictors: Surface_m2, Lake_ha, Usage, DistanceToCentre_m
(f) Predictors: Surface_m2, Lake_ha, Usage, DistanceToCentre_m, Dist_Railway_m
(g) Dependent Variable: EuclideanDist
(h) Linear Regression through the Origin

c) ANOVA(g,h)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	423.171	1	423.171	301.857	.000(a)
	Residual	98.132	70	1.402		
	Total	521.303(b)	71			
2	Regression	460.907	2	230.454	263.285	.000(c)
	Residual	60.396	69	.875		
	Total	521.303(b)	71			
3	Regression	483.925	3	161.308	293.457	.000(d)
	Residual	37.378	68	.550		
	Total	521.303(b)	71			
4	Regression	488.513	4	122.128	249.543	.000(e)
	Residual	32.790	67	.489		
	Total	521.303(b)	71			
5	Regression	490.734	5	98.147	211.906	.000(f)
	Residual	30.569	66	.463		
	Total	521.303(b)	71			

(a) Predictors: Surface_m2
(b) This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.
(c) Predictors: Surface_m2, Lake_ha
(d) Predictors: Surface_m2, Lake_ha, Usage
(e) Predictors: Surface_m2, Lake_ha, Usage, DistanceToCentre_m
(f) Predictors: Surface_m2, Lake_ha, Usage, DistanceToCentre_m, Dist_Railway_m
(g) Dependent Variable: EuclideanDist
(h) Linear Regression through the Origin

d) Coefficients(a,b)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	Surface_m2	.027	.002	.901	17.374	.000	1.000	1.000
2	Surface_m2	.023	.001	.740	15.503	.000	.737	1.357
	Lake_ha	.001	.000	.313	6.566	.000	.737	1.357
3	Surface_m2	.020	.001	.651	16.149	.000	.650	1.539
	Lake_ha	.001	.000	.322	8.502	.000	.736	1.359
	Usage	.217	.034	.227	6.471	.000	.858	1.166
4	Surface_m2	.015	.002	.489	7.522	.000	.222	4.503
	Lake_ha	.001	.000	.306	8.478	.000	.721	1.388
	Usage	.213	.032	.223	6.731	.000	.856	1.168
	DistanceToCentre_m	4.62E-005	.000	.196	3.062	.003	.230	4.351
5	Surface_m2	.012	.002	.407	5.552	.000	.165	6.064
	Lake_ha	.001	.000	.312	8.864	.000	.716	1.397
	Usage	.211	.031	.221	6.851	.000	.856	1.169
	DistanceToCentre_m	3.87E-005	.000	.164	2.568	.013	.218	4.587
	Dist_Railway_m	.000	.000	.126	2.190	.032	.267	3.740

(a) Dependent Variable: EuclideanDist
(b) Linear Regression through the Origin

8.18 MRE - Five Predictors Simultaneously (LN_ED, no Constant)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics(b)

	Mean(a)	Root Mean Square	N
LN_EuclideanDist	.7866	.89499	71
Surface_m2	86.3683	88.96258	71
Lake_ha	792.1803	1463.03461	71
Usage	.9925	2.83338	71
DistanceToCentre_m	9929.4531	11471.0866	71
Dist_Railway_m	884.1523	1036.78556	71

(a) The observed mean is printed
(b) Coefficients have been calculated through the origin.

b) Model Summary(c,d)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.946(b)	.895	.887	.30111	1.777

(a) For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared to R Square for models which include an intercept.
(b) Predictors: Dist_Railway_m, Usage, Lake_ha, DistanceToCentre_m, Surface_m2
(c) Dependent Variable: LN_EuclideanDist
(d) Linear Regression through the Origin

c) ANOVA(c,d)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50.888	5	10.178	112.254	.000(a)
	Residual	5.984	66	.091		
	Total	56.872(b)	71			

(a) Predictors: Dist_Railway_m, Usage, Lake_ha, DistanceToCentre_m, Surface_m2
(b) This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.
(c) Dependent Variable: LN_EuclideanDist
(d) Linear Regression through the Origin

d) Coefficients(a,b)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	Surface_m2	.004	.001	.423	4.297	.000	.165	6.064
	Lake_ha	.000	.000	.286	6.065	.000	.716	1.397
	Usage	.057	.014	.179	4.155	.000	.856	1.169
	DistanceToCentre_m	1.45E-005	.000	.185	2.166	.034	.218	4.587
	Dist_Railway_m	9.15E-005	.000	.106	1.373	.174	.267	3.740

(a) Dependent Variable: LN_EuclideanDist
(b) Linear Regression through the Origin

8.19 MRE - Four Predictors Simultaneously (ED, no Constant)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics(b)

	Mean(a)	Root Mean Square	N
EuclideanDist	2.4235	2.70967	71
Surface_m2	86.3683	88.96258	71
Lake_ha	792.1803	1463.03461	71
Usage	.9925	2.83338	71
DistanceToCentre_m	9929.4531	11471.08662	71

(a) The observed mean is printed
(b) Coefficients have been calculated through the origin.

b) Model Summary(c,d)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.968(b)	.937	.933	.69958	1.759

(a) For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared to R Square for models which include an intercept,
(b) Predictors: DistanceToCentre_m, Usage, Lake_ha, Surface_m2
(c) Dependent Variable: EuclideanDist
(d) Linear Regression through the Origin

c) ANOVA(c,d)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	488.513	4	122.128	249.543	.000(a)
	Residual	32.790	67	.489		
	Total	521.303(b)	71			

(a) Predictors: DistanceToCentre_m, Usage, Lake_ha, Surface_m2
(b) This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin,
(c) Dependent Variable: EuclideanDist,
(d) Linear Regression through the Origin

d) Coefficients(a,b)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	Surface_m2	.015	.002	.489	7.522	.000	.222	4.503
	Lake_ha	.001	.000	.306	8.478	.000	.721	1.388
	Usage	.213	.032	.223	6.731	.000	.856	1.168
	DistanceToCentre_m	4.62E-005	.000	.196	3.062	.003	.230	4.351

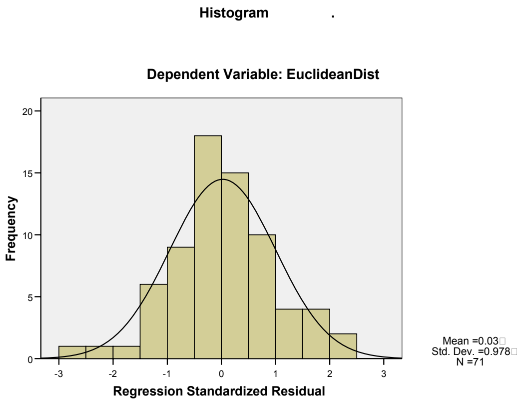
(a) Dependent Variable: EuclideanDist
(b) Linear Regression through the Origin

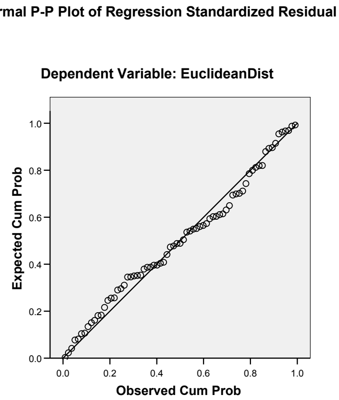
e) Residuals Statistics(a,b)

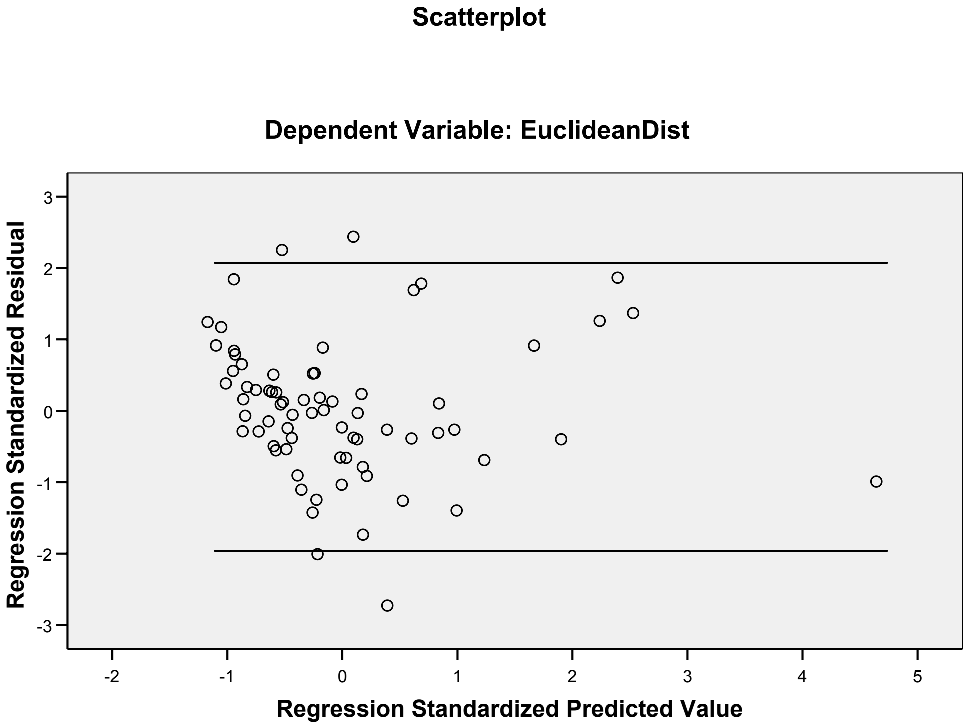
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.1729	7.2907	2.4060	1.05234	71
Residual	-1.90731	1.70623	.01750	.68419	71
Std. Predicted Value	-1.172	4.642	.000	1.000	71
Std. Residual	-2.726	2.439	.025	.978	71

(a) Dependent Variable: EuclideanDist
(b) Linear Regression through the Origin

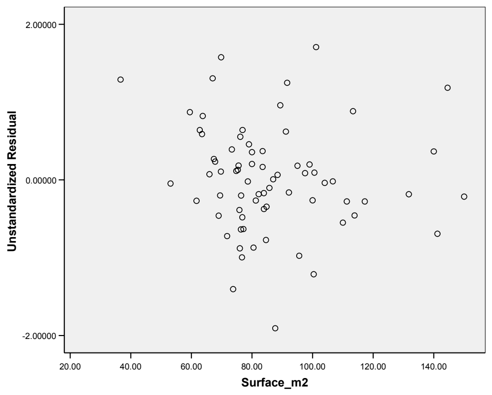
f) Graphical Residual Analysis: Histogram, P-P Plot, Scatterplot

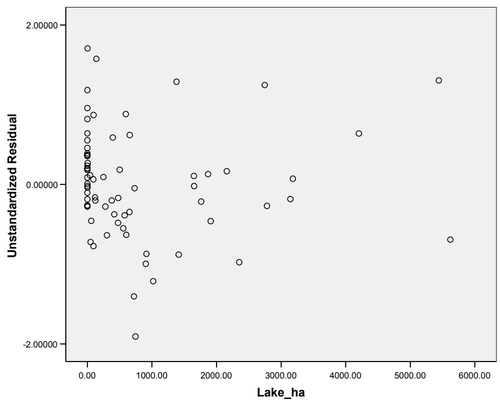


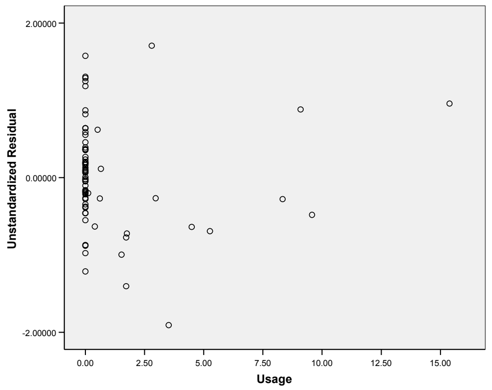


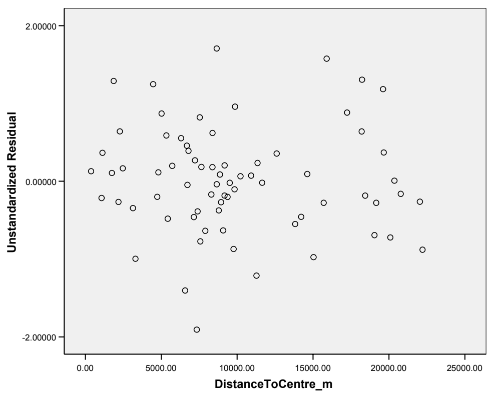


g) Scatterplot Unstandardised Residuals with Variables









h) Statistical Assumptions for Multivariate Regression

1) Linear regression model

Simple inspection of scatterplots is a common, if non-statistical, method of determining if nonlinearity exists in a relationship. According to Garson (2007, p. 10) ‘in regression, as a rule of thumb, an indicator of possible nonlinearity is when the standard deviation of the residuals exceeds the standard deviation of the dependent.’ In this model the standard deviation of the residual (0.684) is smaller than the standard deviation of the dependent variable (1.052) and from the scatterplot of residuals against predicted values, it can be seen that there is no clear relationship between the residuals and the predicted values, consistent with the assumption of linearity.

2) X values are fixed in repeated sampling

According to Hill et al. (2001, p. 283) also if this assumption should be violated ‘... the properties of the least squares estimator (LSE) would still hold.’ This model is based on LSE method thus this assumption appears not to be violated.

3) Zero mean value of residuals

This assumption is met according to the mathematical definition of the least squares regression and confirmed by the approximately zero mean of the residual statistics.

4) Homoscedasticity or equal variance of residuals

According to the scatterplot, there is no evidence of a violation of the homoscedasticity assumption. In fact, the distribution of the variance of residuals seems to be constant.

5) No autocorrelation between the residuals

The Durbin-Watson coefficient of this model suggests that there is no problem. Garson (2007, p. 10) defines the result of this test between 1.5 and 2.5 for independent observations. In this model, the test result of 1.759 lies within the band width indicated.

6) Zero covariance between residual and predictor -> independence of residuals

This assumption is automatically fulfilled if x variable is non-random and assumption 3) “Zero mean value of residuals” holds (Gujarati, 2003). In addition, according to the scatterplots, there is no evidence of violation.

7) Number of observations n must be greater than number of parameters to be estimated

With a proportion of 17.7 (71/4) observations for each predictor, the assumption is met. This assumption is met also under more restrictive researchers. According to Coakes and Steed (2007, p. 136) ‘the minimum requirement is to have at least five times more cases than independent variables.’ According to Kleinbaum et al. (1998) a larger number of independent observations are needed to estimate reliably a larger number of regression coefficients. They affirm that the most basic constraint is that the error degrees of freedom must be positive (d. f. error = $n - k - 1 > 0$) which is equivalent to the constraint $n > k + 1$. Where “n” is the number of observation (71) and “k” is the number of predictors (4), giving $k + 1$ (5) regression coefficients including the intercept.

8) Variability in predictors (X) values

The analysis of the PFs-ZH-Dataset demonstrates no violation of this assumption.

9) Regression model is correctly specified

According to the described methodology for choosing the statistical model that can be used to analyse a correlational research, this assumption is met.

10) There is no perfect multicollinearity

Using the VIF-test the multicollinearity can be tested. Studenmund (2006) defines that a severe multicollinearity is given if the VIF-test > 5. According to the collinearity statistics for this model all the VIF-values are under 5, this means that there is no evidence of violation of this assumption for the FCP-Model estimated.

11) The residuals (error terms) should be normally distributed

Histogram and P-P Plot for this model do not give indication of violation of this assumption.

12) Multivariate Outliers

The Mahalanobis distance is used to test this multivariate model for outliers. An examination of the Mahalanobis distance values indicates that there is one multivariate outlier among the independent variables. In fact, its value is greater than to the critical chi-square value of 18.467 (df 4) at an alpha level of 0.001. According to Hair et al. (2006) outliers do not have to be categorically removed if no plausible reason justifies the removal. This outlier would suggest that there is a slight problem, but because all the other diagnostic data are satisfactory and all others tests seem to confirm the validity of the model, this outlier remains in the dataset.

8.20 MRE - Curve Fitting of Variable (ED, no Constant)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics(b)

	Mean(a)	Root Mean Square	N
EuclideanDist	2.4235	2.70967	71
CF_Surface_m2	2.4179	2.45987	71
Usage	.9925	2.83338	71
CF_Lake_ha	1.2999	1.97906	71
DistanceToCentre_m	9929.4531	11471.0866	71

(a) The observed mean is printed
(b) Coefficients have been calculated through the origin.

b) Model Summary(c,d)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.958(b)	.918	.914	.79637	1.692

(a) For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared to R Square for models which include an intercept.
(b) Predictors: DistanceToCentre_m, Usage, CF_Lake_ha, CF_Surface_m2
(c) Dependent Variable: EuclideanDist
(d) Linear Regression through the Origin

c) ANOVA(c,d)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	478.811	4	119.703	188.743	.000(a)
	Residual	42.492	67	.634		
	Total	521.303(b)	71			

(a) Predictors: DistanceToCentre_m, Usage, CF_Lake_ha, CF_Surface_m2
(b) This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.
(c) Dependent Variable: EuclideanDist
(d) Linear Regression through the Origin

d) Coefficients(a,b)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	CF_Surface_m2	.500	.084	.454	5.943	.000	.208	4.800
	Usage	.213	.036	.223	5.954	.000	.867	1.153
	CF_Lake_ha	.329	.063	.240	5.192	.000	.567	1.763
	DistanceToCentre_m	5.90E-005	.000	.250	3.586	.001	.251	3.982

(a) Dependent Variable: EuclideanDist
(b) Linear Regression through the Origin

8.21 Cubic Curve Fitting -> EuclideanDist = f (Surface_m2)

Following statistics are developed from field data using SPSS.

Curve Fitting (CF)

$$CF\_Surface = 0.084*(Surface\_m2) – 0.01*(Surface\_m2)^2 + 4.80E\text{-}006 *(Surface\_m2)^3$$

a) Model Summary(a)

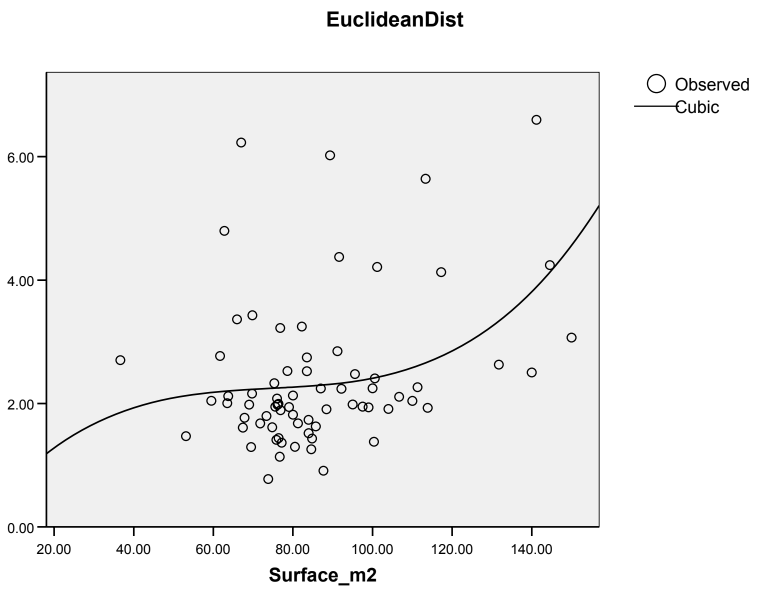
R	R Square	Adjusted R Square	Std. Error of the Estimate
.908	.824	.816	1.161

The independent variable is Surface_m2.
(a) The equation was estimated without the constant term.

b) Coefficients

	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error	Beta		
Surface_m2	.084	.026	2.743	3.181	.002
Surface_m2 ** 2	-.001	.001	-3.553	-2.076	.042
Surface_m2 ** 3	4.80E-006	.000	1.808	1.979	.052

c) Curve Fitting



8.22 Cubic Curve Fitting -> EuclideanDist = f (Lake_ha)

Following statistics are developed from field data using SPSS.

Curve Fitting (CF)

CF_Lake_ha = 0.004*(Lake_ha) - 1.35E-006*(Lake_ha)2 + 1.67E-010*(Lake_ha)3

a) Model Summary(a)

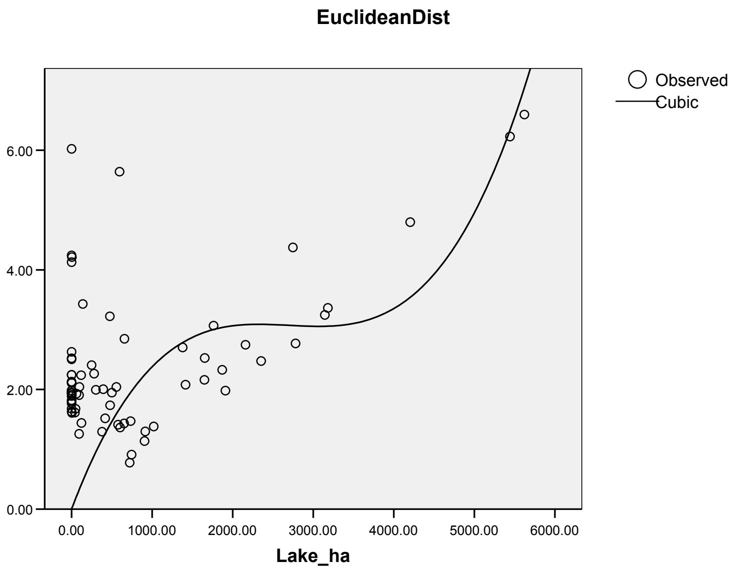
R	R Square	Adjusted R Square	Std. Error of the Estimate
.730	.533	.513	1.891

The independent variable is Lake_ha.
(a) The equation was estimated without the constant term.

b) Coefficients

	Unstandardised Coefficients		Standardised Coefficients	t	Sig.
	B	Std. Error	Beta		
Lake_ha	.004	.001	1.922	4.212	.000
Lake_ha ** 2	-1.35E-006	.000	-3.019	-2.517	.014
Lake_ha ** 3	1.67E-010	.000	1.880	.	.

c) Curve Fitting



8.23 MRE - Transformations of Variables (ED, no Constant)

Following statistics are developed from field data using SPSS.

a) Descriptive Statistics(b)

	Mean(a)	Root Mean Square	N
EuclideanDist	2.4235	2.70967	71
Lake_ha	792.1803	1463.03461	71
Usage	.9925	2.83338	71
DistanceToCentre_m	9929.4531	11471.08662	71
TR_LN_Surface_m2_Norm	4.4299	4.43633	71

(a) The observed mean is printed
(b) Coefficients have been calculated through the origin.

b) Model Summary(c,d)

Model	R	R Square(a)	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.966(b)	.933	.929	.72132	1.641

(a) For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This cannot be compared to R Square for models which include an intercept.
(b) Predictors: TR_LN_Surface_m2_Norm, Usage, Lake_ha, DistanceToCentre_m
(c) Dependent Variable: EuclideanDist
(d) Linear Regression through the Origin

c) ANOVA(c,d)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	486.443	4	121.611	233.734	.000(a)
	Residual	34.860	67	.520		
	Total	521.303(b)	71			

(a) Predictors: TR_LN_Surface_m2_Norm, Usage, Lake_ha, DistanceToCentre_m
(b) This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.
(c) Dependent Variable: EuclideanDist
(d) Linear Regression through the Origin

d) Coefficients(a,b)

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tol.	VIF
1	Lake_ha	.001	.000	.293	7.783	.000	.705	1.419
	Usage	.223	.032	.233	6.872	.000	.866	1.155
	DistanceToCentre_m	5.13E-005	.000	.217	3.332	.001	.235	4.252
	TR_LN_Surface_m2_Norm	.287	.041	.469	7.017	.000	.223	4.481

(a) Dependent Variable: EuclideanDist
(b) Linear Regression through the Origin

8.24 LN Transformation TR_LN_Lake_ha = LN(Lake_ha)

Following statistics are developed from field data using SPSS.

Transformation (TR)
TR_LN_Surface_m2 = Ln (Surface_m2)

a) Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Surface_m2	71	100.0%	0	.0%	71	100.0%
TR_LN_Surface_m2	71	100.0%	0	.0%	71	100.0%

b) Tests of Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Surface_m2	.135	71	.003	.920	71	.000
TR_LN_Surface_m2	.088	71	.200(*)	.961	71	.027

(*) This is a lower bound of the true significance.
(a) Lilliefors Significance Correction

The variable “Surface_m2” seems to have a not normal distribution. The new variable “TR_LN_Surface_m2” seems to have a normal distribution in accordance to the tests of normality (KS, SW).

8.25 Descriptive Statistic for Variable "NrRooms", N=17,060

Following statistics are developed from field data using SPSS.

a) Descriptive Statistic

			Statistic	Std. Error
NrRooms	Mean		3.1935	.00815
	95% Confidence Interval for Mean	Lower Bound	3.1775	
		Upper Bound	3.2095	
	5% Trimmed Mean		3.2021	
	Median		3.0000	
	Variance		1.133	
	Std. Deviation		1.06444	
	Minimum		1.00	
	Maximum		8.00	
	Range		7.00	
	Interquartile Range		1.00	
	Skewness		-.132	.019
	Kurtosis		-.015	.038

b) Explorative Statistic

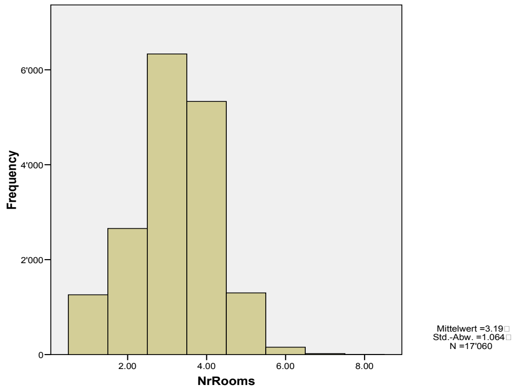
Case Processing Summary

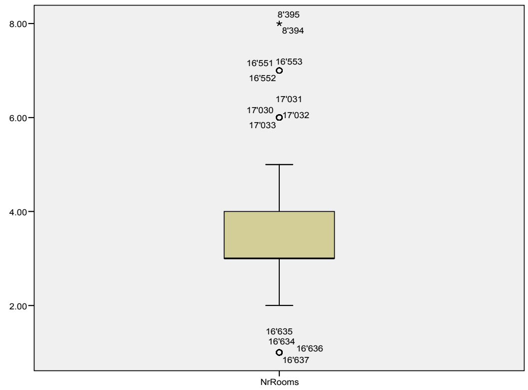
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
NrRooms	17060	100.0%	0	.0%	17060	100.0%

Extreme Values

			Case Number	Value
NrRooms	Highest	1	8394	8.00
		2	8395	8.00
		3	543	7.00
		4	2849	7.00
		5	2850	7.00
	Lowest	1	16637	1.00
		2	16636	1.00
		3	16635	1.00
		4	16634	1.00
		5	16633	1.00

c) Histogram and Boxplot





8.26 Descriptive Statistic for Variable "Surface", N=16,215

Following statistics are developed from field data using SPSS.

a) Descriptive Statistic

			Statistic	Std. Error
Surface	Mean		79.2070	.21056
	95% Confidence Interval for Mean	Lower Bound	78.7943	
		Upper Bound	79.6198	
	5% Trimmed Mean		78.5491	
	Median		77.0000	
	Variance		718.894	
	Std. Deviation		26.8122	
	Minimum		12.00	
	Maximum		360.00	
	Range		348.00	
	Interquartile Range		35.00	
	Skewness		.621	.019
	Kurtosis		2.255	.038

b) Explorative Statistic

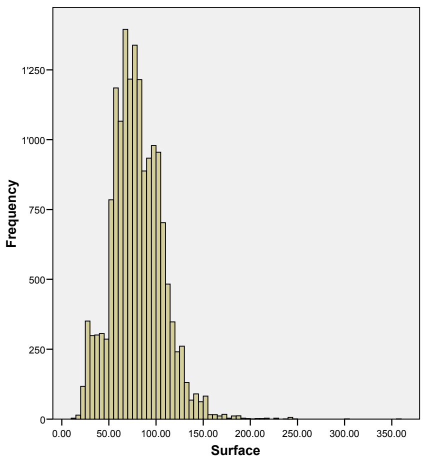
Case Processing Summary

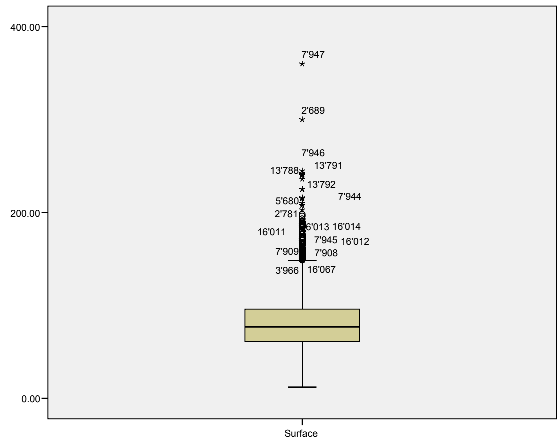
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Surface	16215	100.0%	0	.0%	16215	100.0%

Extreme Values

			Case Number	Value
Surface	Highest	1	7947	360.00
		2	2689	300.00
		3	5682	245.00
		4	2688	242.00
		5	13789	241.00
	Lowest	1	8636	12.00
		2	8637	14.00
		3	6262	14.00
		4	8639	16.00
		5	8638	16.00

c) Histogram and Boxplot





8.27 Descriptive Statistic for Variable "Volume", N=15,836

Following statistics are developed from field data using SPSS.

a) Descriptive Statistic

			Statistic	Std. Error
Volume	Mean		9460.2390	86.9362
	95% Confidence Interval for Mean	Lower Bound	9289.8342	
		Upper Bound	9630.6439	
	5% Trimmed Mean		7702.4142	
	Median		6279.0000	
	Variance		119686938.49	
	Std. Deviation		10940.15258	
	Minimum		402.00	
	Maximum		82430.0	
	Range		82028.00	
	Interquartile Range		7247.00	
	Skewness		3.622	.019
	Kurtosis		16.266	.039

b) Explorative Statistic

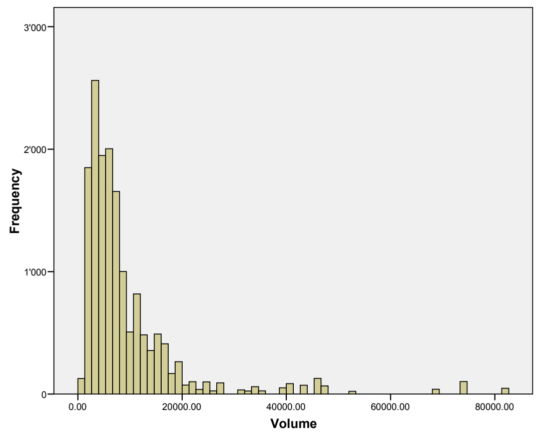
Case Processing Summary

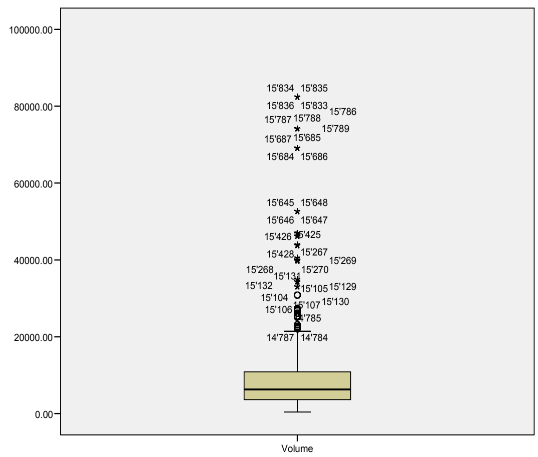
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Volume	15836	100.0%	0	.0%	15836	100.0%

Extreme Values

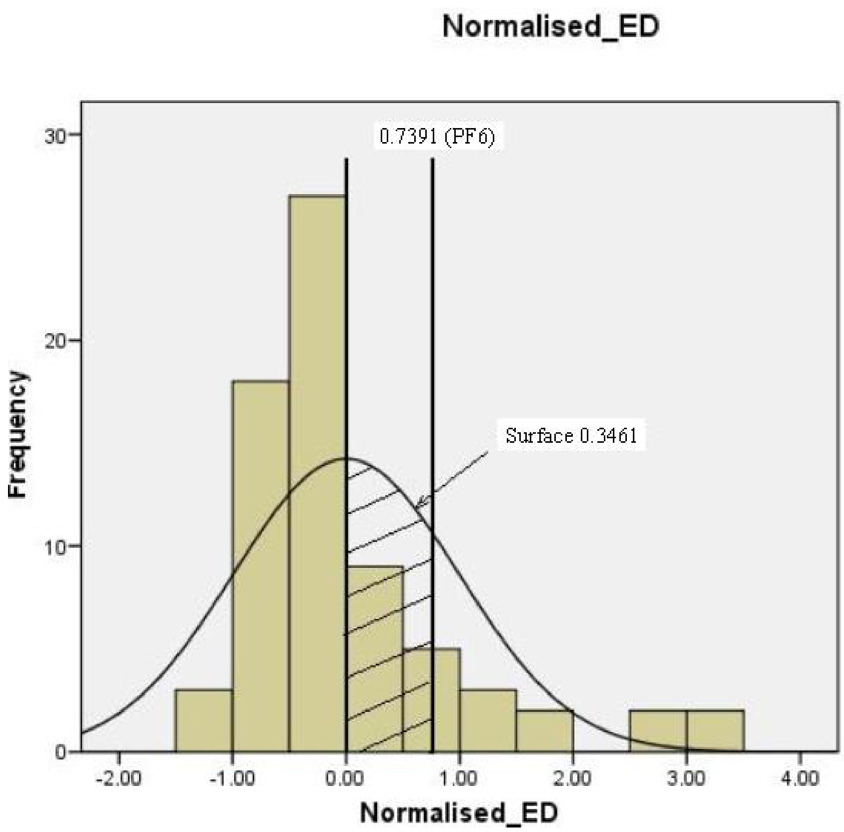
			Case Number	Value
Volume	Highest	1	15790	82430.0
		2	15791	82430.0
		3	15792	82430.0
		4	15793	82430.0
		5	15794	82430.0
	Lowest	1	1	402.00
		2	2	410.00
		3	3	464.00
		4	5	466.00
		5	4	466.00

c) Histogram and Boxplot





8.28 Histogram Normalised Euclidean Distance



(Source: Developed for this research)