Year : 2019

ESSAYS ON LONG-TERM CARE AND INSURANCE MANAGEMENT

Rudnytskyi Iegor

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

ESSAYS ON LONG-TERM CARE AND INSURANCE MANAGEMENT

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Actuarielles

par

Iegor RUDNYTSKYI

Directeur de thèse
Prof. Joël Wagner

Jury

Prof. Christian Zehnder, Président
Prof. François Dufresne, expert interne
Prof. Frédéric Planchet, expert externe

LAUSANNE
2019

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

ESSAYS ON LONG-TERM CARE AND INSURANCE MANAGEMENT

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Actuarielles

par

Iegor RUDNYTSKYI

Directeur de thèse
Prof. Joël Wagner

Jury

Prof. Christian Zehnder, Président
Prof. François Dufresne, expert interne
Prof. Frédéric Planchet, expert externe

LAUSANNE
2019

Unil
UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Iegor RUDNYTSKYI, titulaire d'un bachelor en mathématiques appliquées de National Technical University «Kharkiv Polytechnic Institute», d'un master en mathématiques appliquées de V.N. Karazin Kharkiv National University, ainsi que d'un master en sciences actuarielles de l'Université de Lausanne, en vue de l'obtention du grade de docteur ès Sciences actuarielles.

La thèse est intitulée :

ESSAYS ON LONG-TERM CARE AND INSURANCE MANAGEMENT

Lausanne, le 11 juin 2019

Le doyen

Jean-Philippe Bonardi

HEC Lausanne
Le Décanat
Tél. +41 21 692 33 40 | Fax +41 21 692 33 05
www.hec.unil.ch | hecdoyen@unil.ch

Members of the Thesis Committee

Professor Joël Wagner: Thesis supervisor, Full Professor at the Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne and Swiss Finance Institute, Switzerland.

Professor François Dufresne: Internal expert, Full Professor at the Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland.

Professor Frédéric Planchet: External expert, Professor at the Institut de Science Financière et d'Assurances (ISFA), University Claude Bernard Lyon 1, France.

Professor Christian Zehnder: Jury president, Full Professor at the Department of Organizational Behavior, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland.

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Iegor RUDNYTSKYI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _______________ Date: 4 June 2019

Prof. Joël WAGNER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Iegor RUDNYTSKYI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: F. Dufresne Date: 4 June 2019

Prof. François DUFRESNE
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Iegor RUDNYTSKYI

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________ Date: 05/06/2019

Prof. Frédéric PLANCHET
External member of the doctoral committee

Declaration of Authorship

I, Iegor Rudnytskyi, declare that this thesis titled, ESSAYS ON LONG-TERM CARE AND INSURANCE MANAGEMENT and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Parts of this thesis appeared in the publications that are indicated in the relevant chapters.

Рудницький

Signature

2019 June 6

Date

Acknowledgements

It would not have been possible to complete this thesis without the valuable support and help of many people, to whom I would like to express my gratitude below.

First off, I would like to thank my supervisor Professor Joël Wagner for giving me the opportunity to implement this thesis by navigating me in the academic world. It is hardly possible to estimate his invaluable contribution, precious input and strong support. It has been many lessons I learned (both academic and non-academic) under his extensive guidance, for which I am very grateful.

I would like to thank my thesis committee members, Professor François Dufresne and Professor Frédéric Planchet, for their precious comments, suggestions and overall feedback. The quality of this thesis was significantly improved because of the time they devoted to read it in depth, which is much appreciated.

My stay at the Department of Actuarial Science was remarkably pleasant. I was very happy to have numerous intellectual discussion with our professors that contributed to the shape of my thinking. I want to thank Hansjörg Albrecher, Séverine Arnold, François Dufresne, Enkelejd Hashorva, Philippe Maeder and Hans-Ulrich Gerber. A warm thank you is also owed to Catherine Lombard Vonlanthen. Further, I do not dare to call other members of DSA as colleagues but rather friends. Many thanks to (in no particular order) to Héloïse, Eleni, Viktoriya, Maissa, José, Léonard, Martin, William, Michel, Charbel, Grigori, Guillem, Anca, Veronika, Nikolai, Yves, Dina, Arian, Philipp, Youssouf, Gildas, Dalit, Liselotte, Jevgenijs, Long and Peng.

I would also like to express my gratitude to Valérie Chavez and Marc-Olivier Boldi for giving me the opportunity to work as an assistant in the Operational Research department. Many thanks go to my colleagues and friends from the other departments: Gabriela, Ina, Nataliya, Oksana, Anastasia, Sarah, Anna, Maximillian, Fabien, Thibault and Samuel.

I am grateful to my friends for the time we spent together: Dasha, Sasha, Anette, Oriol, Olga, Rita, Kirill, Zhenya, Nastya, Kasim, Yana, Daria and Viktor Yurovsky.

Last but not least, I do not have enough words to express how grateful I am to my parents, my fiancée Lika, my brother Anton and our cat Churchill. They supported me in any possible way.

Summary

This thesis is composed of two parts. The first part encompasses two essays on long-term care (LTC) in Europe and Switzerland. LTC can be defined as help provided to elderly in needs with activities of daily living (ADL) that can be delivered as informal care by persons from the social network or as formal care by professional caregivers. The first paper focuses on individuals' characteristics that influence the probability to report limitations in ADLs and the probability to report the receipt of formal care in Europe. We find a significant effect of social, demographic and medical factors on both quantities of interest. Our results highlight the lower effect of cancer when compared to other diseases. Further, we observe a decrease of formal care usage when the partner is present in the household. This effect is stronger for domestic tasks and in case of male respondents. In addition, elderly in countries with family care LTC schemes report less formal care than those in countries with other schemes. The second essay extends the measure of dependence by considering instrumental ADLs and functional limitations. Thereby the focus is on identifying the relationship between formal and informal care, that is, whether this link is complementary or substitutional. By using a Structural Equation Modeling approach and data from a Swiss survey, the link is found to be both of complementary and substitutional nature. The second part of the thesis is dedicated to insurance management. We study the impact of natural and man-made catastrophes on the valuation of insurance companies. Furthermore, we analyze the relationship between the effect on the valuation and companies' characteristics, namely the market capitalization, the subsector, the revenues and the geographical origin. We do not find any clear pattern of the stock price behavior for any type of the considered catastrophes. Despite the fact, that we find no significance of the geographical origin coefficient in the regression, we observe that North American companies are more influenced by local events, such as hurricanes on US territory and experience almost no effect from external ones. In contrast, European companies respond to all events, including international ones. Further, reinsurance companies are found to be more sensitive than property and casualty (P&C) companies. Finally, in the last essay, we study the properties of stochastic programming for the asset allocation in a framework of Swiss pension funds under solvency constraints. We empirically study the convergence of the initial asset allocation with respect to the quality of the approximation of the stochastic returns. We observe results for the probability of deficit as well as for the expected value of the deficit given shortage. Further, we test the sensitivity of the above-mentioned characteristics to changes in internal model parameters. Finally, the effect of misestimating the stocks' volatility and the bonds' expected return on the initial asset allocation is examined.

Résumé

Cette thèse est composée de deux parties. La première partie comprend deux essais sur les soins de longue durée en Europe et en Suisse. Les soins de longue durée peuvent être définis comme l'aide fournie aux personnes âgées ayant des difficultés avec les activités de la vie quotidienne (AVQ) et peuvent être dispensés sous forme de soins informels par des personnes de leur entourage ou sous forme de soins formels dispensés par des aidants professionnels. Le premier essai porte sur les caractéristiques des individus qui influencent la probabilité de rapporter des limitations dans les AVQ et sur la probabilité de recourir à des soins formels en Europe. Nous constatons un effet significatif des facteurs sociaux, démographiques et médicaux sur les deux quantités d'intérêt. Nos résultats mettent en évidence un effet moindre du cancer par rapport à d'autres maladies. De plus, nous observons une diminution de l'utilisation des soins formels lorsque le partenaire est présent dans le ménage. Cet effet est plus fort pour les tâches domestiques et dans le cas des sujets masculins. En outre, les personnes âgées des pays où les régimes de soins de longue durée sont basés sur l'aide de la famille font état de moins de soins formels que celles des pays dotés d'autres régimes. Le deuxième essai se concentre sur la mesure de la dépendance en prenant en compte les activités instrumentales de la vie quotidienne (AIVQ) et les limitations fonctionnelles. Ainsi, l'accent est mis sur l'identification de la relation entre les soins formels et informels, c'est-à-dire si ce lien est complémentaire ou substitutionnel. En utilisant une approche de modélisation par équation structurelle et des données provenant d'une enquête suisse, le lien s'avère être à la fois complémentaire et substitutif. La deuxième partie de la thèse est consacrée à la gestion des assurances. Nous étudions l'impact des catastrophes naturelles et d'origine humaine sur l'évaluation des sociétés d'assurance. En outre, nous analysons la relation entre les caractéristiques des sociétés et l'effet sur leur évaluation, à savoir la capitalisation boursiére, le sous-secteur, les revenus et l'origine géographique. Pour tous les événements catastrophiques considérés, nous ne trouvons pas de tendance claire dans le comportement du cours de l'action. Malgré le fait que le coefficient de régression correspondant à l'origine géographique n'est pas significatif, nous observons que les entreprises nord-américaines sont davantage influences par des événements locaux, tels que des ouragans sur le territoire américain, et pratiquement pas affectées en cas de catastrophes en dehors du territoire. En revanche, les entreprises européennes sont sensibles à tous les événements, y compris internationaux. En outre, les sociétés de réassurance se révèlent plus sensibles que les sociétés d'assurance Incendie, Accidents et Risques Divers (IARD). Enfin, dans le dernier essai, nous étudions les propriétés de la programmation stochastique dans le cadre de la répartition des actifs pour des fonds de pension suisses soumis à des contraintes de solvabilité. Nous étudions de manière empirique la convergence de la répartition initiale des actifs par rapport à la qualité de l'approximation des rendements stochastiques. Nous observons des résultats pour la probabilité de ruine ainsi que pour la valeur attendue du déficit en cas de défaut. De plus, nous testons la sensibilité des caractéristiques susmentionnées en modifiant différents paramètres internes du modèle. Enfin, les effets d'une mauvaise estimation de la volatilité des actions et du rendement attendu des obligations sur la répartition initiale des actifs sont examinés.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is composed of two parts. The first part deals with the topic of long-term care and integrates two essays on the characterization of care for elderly in Europe and Switzerland. In the second part, two essays focus on insurance management: on the one hand, an event study analyses the impact of catastrophe events on the market valuation of insurance companies, and, on the other hand, a stochastic programming note studies the process of finding the optimal asset allocation in pension funds.

Long-term care

The population is aging in many industrialized countries in the world and the increase in the number of elderly people triggers a higher demand for long-term care (LTC). LTC can be described as measures, assistance and services provided to elderly who cannot take care for themselves independently. The demand for such care is driven by many aspects related to the health condition as well as demographic and social factors, for example. Further, LTC can be delivered as informal care by persons from the social environment, i.e., partners, children, friends or neighbors, of a dependent elderly. Further, formal LTC embodies services and assistance in personal care and domestic tasks by specifically trained and qualified persons. Formal LTC is associated with higher expenditures and raises financial challenges on the health policies. Informal care can be an important alternative to more costly formal care. However, the contribution by households can be very high.

LTC is associated with limitations in activities of daily living (ADL), including dressing, bathing, getting in and out of bed, toileting, walking and eating. In Chapter 1 we aim to derive the drivers affecting the probability of reporting limitations in ADL and the probability of demanding formal LTC. As formal care, we consider such types of at-home care as help provided with domestic tasks, personal care, meals-on-wheels and other activities as well as formal care provided in nursing homes. By using the most recent wave of a cross-national European survey on individuals aged over 50 years (SHARE, wave 6), we develop econometric models for identifying the effect of demographic, social and medical factors on ADL limitations and formal LTC along five conjectures. In our study, we focus on individuals aged over 65 years from selected European countries. These countries are classified according to the LTC scheme they use. On the one hand, we analyze functional limitations and we find that characteristics such as the age, the gender, the wealth status and the education level influence the probability to report limitations. Further, while we find that pathologies significantly increase the probability to become dependent in general, the effect of cancer is lower. On the other hand, we find again an influence of the demographic and social factors on the probability to use formal LTC. We emphasize on the decrease in the probability due to the presence of the partner in the household, in particular for housekeeping tasks. This is less the case for help related with personal care. In addition, we note that pathologies such as cancer have no influence on the probability to report formal LTC while others like mental and Parkinson diseases highly increase it. We find that elderly living in countries with LTC family care schemes report less formal care than in others. This indicates the importance of LTC policies. Finally, we validate the robustness of our results by applying the models to data from earlier waves of the survey.

In Chapter 2 we extent the measurement of dependence of elderly by using additional scales of instrumental ADL and functional limitations. Once an elderly fails to perform these activities independently, he or she requires special assistance. The aim of this research is to study individual characteristics that can help to determine the demand of LTC and define a relation between formal and informal care. Our study is based on data from the Swiss Health Survey and focuses again on respondents aged over 65 years. We develop a statistical model using the structural equation modeling technique that allows representing the dependence concept as a latent variable. This hidden dependence variable combines indices linked to limitations in ADL, instrumental ADL and functional limitations. Accounting for causality links between covariates enables us to include the indirect effect of pathologies on the receipt of LTC mediated via dependence. Furthermore, in our model we do not assume a causal relationship between formal and informal care. From our results, we observe a significant impact of pathologies as well as of the social environment on the demand for LTC. The relationship between formal and informal care is found to be both of complementary and substitutional nature.

Catastrophe events and market valuation

In the further chapters of this thesis, we cover other topics in insurance management that are unrelated to LTC. In Chapter 3, we apply an event study methodology to examine the impact of a selected number of major catastrophe events on the stock valuation of the 87 largest listed non-life insurance companies worldwide. First, we review six parametric and six nonparametric statistical tests from event study and validate the performance of these tests based on the well-known event of the 2001 9/11 terrorist attacks. Then, we aim to study if the responses from insurance companies' stock prices depend on the events' nature. In our analysis, we consider airline crashes, earthquakes, hurricanes and winter storms. For each event, we study the significance of its impact on a set of insurers. Depending on the type of the event, the produced effect, its significance and its extent are different. Furthermore, we analyze the relation between the caused valuation effects and the companies' characteristics such as the market capitalization, the revenue, the relevant sectors of insurance, the geographical origin and the split of revenues. Thereby, we try to explain which companies' characteristics drive the stock market response. The obtained results are mixed: for some events our research gives indications for the management of insurance companies during catastrophe events while in other cases it shows the limits of applicability of event study analyses.

Asset allocation in pension funds

In Chapter 4, we study the properties of stochastic programming for the asset allocation in a framework of Swiss pension funds under solvency constraints. We develop a simplified yet scalable model using a linear utility function as an objective function. In the framework of our utility function, we let the future liability be deterministic, while we model the random returns of assets by using a vector autoregressive model as the underlying econometric model. Then, we calibrate such model using monthly returns of bonds and stock indices. In our case, the stochastic programming problem cannot be solved exactly, that is why the deterministic equivalent represented by a scenario tree is used. The scenario tree is generated by discretizing the econometric model using the "bracket-mean" method. In this chapter, we empirically study the convergence of the initial asset allocation with respect to the number of scenarios that depends on the discretization parameter. We also observe the results for the probability of deficit as well as for the expected value of the deficit given shortage. Further, we test the sensitivity of the above-mentioned characteristics to changes in internal parameters of the model, namely, planning horizon, target wealth and shortage penalty. Finally, the effect of misestimating the stock volatility and the bonds' expected return on the initial asset allocation is examined.

Chapter 2

On the Characteristics of Reporting ADL Limitations and Formal LTC Usage across Europe

The increase in the proportion of elderly people in most industrialized countries triggers higher demand for long-term care (LTC) associated with limitations in activities of daily living (ADL). The aim of this research is to derive the drivers affecting the probability of reporting limitations in ADL and the probability of demanding formal LTC, e.g., personal care and services in domestic tasks. By using the most recent wave of a cross-national European survey on individuals aged over 50 years (SHARE, wave 6), we develop econometric models for identifying the effect of demographic, social and medical factors on ADL limitations and formal LTC along five conjectures. On the one hand, we analyze functional limitations and we find that characteristics such as the age, the gender, the wealth status and the education level influence the probability to report limitations. Further, while we find that pathologies significantly increase the probability to become dependent in general, the effect of cancer is lower. On the other hand, we find again an influence of the demographic and social factors on the probability to use formal LTC. We emphasize on the decrease in the probability due to the presence of the partner in the household, in particular for housekeeping tasks. This is less the case for help related with personal care. In addition, we note that pathologies such as cancer have no influence on the probability to report formal LTC while others like mental and Parkinson diseases highly increase it. We find that elderly living in countries with LTC family care schemes report less formal care than in others. This indicates the importance of LTC policies. Finally, we validate the robustness of our results by applying the models to data from earlier waves of the survey.

Note: This is joint work with M. Fuino (Department of Actuarial Science, University of Lausanne) and J. Wagner. Both coauthors acknowledge financial support from the Swiss National Science Foundation (grant no. 100018_169662). The authors are thankful for the comments on earlier versions of this manuscript from participants in the following conferences and seminars: European Actuarial Journal Conference 2018 (Leuven, Sept. 2018), L2 Seminar (Lyon, Jan. 2019), Deutscher Verein für Versicherungswissenschaft (Berlin, Mar. 2019) and International Congress on Insurance: Mathematics and Economics (Munich, Jul. 2019). The paper has been submitted to the *Journal of Risk and Insurance*.

2.1 Introduction

Handling the forthcoming high number of elderly and in particular the financing threat and infrastructure needs due to the demand for long-term care (LTC) is at the foreground of many policy debates in Europe (Eurostat, 2017). In this context, more and more developed countries consider LTC as a new social risk (Da Roit et al., 2007). For example, since the eighties, France considers including the dependence of elderly as a fifth risk of the social security. LTC characterizes the help provided to elderly in need of assistance with the activities of daily living (ADL), namely, bathing, dressing, using the toilet, transferring in and out of a bed or a chair, continence and feeding (Katz and Jackson, 1963). Such care is mainly delivered to individuals aged over 65 years (Balia and Brau, 2014) with prevalence rates rising exponentially after the age of 80 years (Fuino and Wagner, 2018b). Typically, two types of care, at-home and institutional care, are distinguished. At-home LTC represents the care an elderly receives in his own house while institutional care refers to the one delivered in a specialized institution. While the first type relates to care received upon request, the latter corresponds to 24-hours supervision in a specialized infrastructure including accommodation and comes at higher costs. Three major questions arise from LTC (Eling and Ghavibazoo, 2018). First, the threat of a financial burden stems from the importance of these costs. As things stand, no state can finance the upcoming burden without modifying current social insurance schemes. Further, the contribution by households is considered as too high in most developed countries (Swiss Re, 2014) and large parts of the population cannot afford it. The second major question stems from understanding how to appropriately let private long-term care insurance (LTCI) take part in the financing (Zhou-Richter et al., 2010; Costa-Font et al., 2019). Most of the reasons yielding the underdevelopment of LTCI come from the individuals' underestimation of LTC risk (Sloan and Norton, 1997), from the discouragement by public policies distributing allowances to the ones without coverage (Courbage and Costa-Font, 2015) and from the help provided by relatives (Costa-Font, 2010). Finally, the demand of LTC is driven by many factors and subject to measurement issues. While LTC help provided by professionals, namely formal care, is statistically measurable, help provided by relatives, i.e. informal care, mostly stays hidden beneath the surface and is not directly observable.

In this paper, we first study the reporting of limitations in ADL among elderly in Europe for understanding how demographic, social and medical factors affect them. Limitations in ADL are essential in measuring LTC needs and many factors can affect them. Usually, at high ages, women have more chances than men to present functional limitations (Fuino and Wagner, 2018b). The socioeconomic status also plays an important role since higher wealth often comes along with better health (Pollack et al., 2007). Further differences in dependence stem from the pathology where, e.g., elderly with diabetes, heart failure and high blood pressure have more chance to require help with ADL than others (Sinclair et al., 2008). In a second part, we investigate on the usage of formal LTC, i.e. help provided by professional caregivers. While functional limitations are an objective measure of elderly dependence, the demand for professional LTC and paid-for services strongly depends on the household composition, financial means and personal believes (Styczynska and Sowa, 2011). In a household, the partner is commonly the first provider of informal care (Pinquart and Sörensen, 2011). In that sense, the number of children can also reduce the demand for formal LTC. As part of their social responsibility, children are often caregivers. However, other important factors such as the distance of the parent's house and the closeness to their parent can play an important role (Courbage and Zweifel, 2011; Steinbeisser et al., 2018). Our empirical approach builds on data coming from thirteen European countries available in the sixth wave of the Survey of Health, Ageing and Retirement in Europe (SHARE, Börsch-Supan et al., 2013). This European cross-national survey on the health of the population aged over 50 years is conducted every two years, and the data contains information on European countries over the period from 2004 onwards.

Our main results quantify the relevance of demographic, social and medical factors on both the probability to report limitations in ADL and formal LTC usage in Europe. From the cross-national study, we find that demographic factors such as the age and the gender have a strong influence on ADL limitations. For ages above 80 years, women are more likely to present limitations than men are. Further, poorer health conditions such as presenting a body mass index away from normal weight as well as being diagnosed with mental, Parkinson, cancer, musculoskeletal system and other physical diseases increase the probability to be dependent. We observe that mental and Parkinson diseases come along with more limitations when compared to other pathologies. Our study also highlights the role of wealth and education in defining elderly health. With higher education, higher wealth and living together with their partner, elderly tend to report fewer ADL limitations. When detailing our results by type of ADL, we observe that the presence of the partner in the household mostly reduces the claiming of difficulties with bathing. Moreover, when considering the probability to report formal LTC, we find that elderly living with their partner require significantly fewer professional services highlighting the importance of informal care. This effect is particularly affecting domestic tasks while for personal care the presence of the partner does not reduce the claiming behavior. Another interesting outcome is related to the country's specific LTC policy. Countries with family care schemes, like Italy or Spain, rely more extensively on family members for delivering LTC and appear in our study to strongly decrease the probability for requesting formal LTC. Finally, by considering pathologies, we note that formal LTC is also more often required in cases where mental and Parkinson diseases are diagnosed. We observe that, while some diseases increase the number of limitations in ADL, they do not necessarily entail more formal LTC.

The remainder of this article is organized as follows: Section 2.2 describes the LTC schemes found in Europe and presents five research conjectures that guide the development of the paper. In Section 2.3, we detail the available data and lay out descriptive statistics on the demographic, social and medical factors that we use. Further, we report on the number of available observations by types of ADL and of formal LTC. In Section 2.4, we introduce the econometric models, present the results of their application on the data and discuss the outcomes and their robustness. Finally, we conclude in Section 2.5.

2.2 LTC policies across Europe and research hypotheses

2.2.1 Landscape of the LTC systems in Europe

LTC refers to the care offered to an elderly in need of assistance in ADL. Depending on the country, the assessment of LTC needs relies on different approaches and recent reforms affect the recognized care levels (Janssen et al., 2016). Nonetheless, they are mostly extensions of the Katz scale (see Katz and Jackson, 1963). In the following, we discuss the LTC schemes found in thirteen European countries that we will later cover with empirical data.

France, Germany and Belgium share common characteristics in their LTC policy. In France, the LTC expenses are financed by both the government and private insurance. The State distributes a benefit named “Allocation Personnalisée d’Autonomie” to elderly over 60 years in need of LTC regardless of their wealth (Courbage and Roudaut, 2008, 2011). The evaluation of the needs is based on a scale combining both instrumental (IADL, including using the phone, using transportation, taking medication and managing money, see for example Brody and Lawton, 2018) and physical ADL (“Autonomie Gérontologie Groupes Iso-Ressources”, see Aguilova et al., 2014). In Germany, LTC is part of the fifth pillar of social security and included in the statutory health insurance system. Since benefits do not cover the full LTC costs, individuals can buy supplementary private insurance providing additional benefits. Similarly to the assessment found in France, elderly are eligible for receiving benefits in case of difficulties with ADL and IADL. For a long time three care levels were distinguished; in 2017, the system has changed and now uses a five-care levels scale (Nadash et al., 2018). Belgium is characterized by a highly developed formal LTC scheme complemented by informal care from the family. The system is universal in the sense that the federal compulsory health insurance provides benefits to the whole population. The benefits are means-tested and calculated upon a combination of ADL and IADL scales. In addition, a separate scheme that pays supplementary cash benefits is available in the Flemish region. In the Flemish care insurance, the LTC assessment includes cognitive and social measurements in addition to ADL and IADL scales (Willemé, 2010).

Considering the Mediterranean countries, we note similarities in the systems of Spain, Italy and Greece. Common characteristics include the scarce development of private health insurance solutions and the important place of the help from family members (Bettio et al., 2006; Jiménez-Martín et al., 2016; Courbage et al., 2018). The LTC system in Spain is composed by a universal allowance scheme covering the whole population and classifying eligible elderly along three dependency levels, mild, moderate and severe, related to the number of limitations in ADL. Only moderately and severely dependent individuals can claim benefits. They can choose between cash or in-kind benefits. While selecting in-kind benefits leads to higher allowances, cash is often preferred since it provides a way to remunerate the significant share of informal care delivered by the family. In addition, local authorities provide further benefits subject to means test. In Italy, the LTC scheme is organized at the State, regional and municipalities levels. The acuity of the dependence is defined along severity scales that differ with the region of residence (Gori, 2008; Tediosi and Gabriele, 2010). The “indemnità di accompagnamento” paid only to severely dependent elderly is the most important cash benefit managed by the Italian social security. Equivalently to Spain, there are other benefits paid by municipalities with eligibility conditions changing by location. The Greek LTC public system provides in-kind and cash benefits to elderly over 65 years showing significant difficulties in performing ADL and IADL as well as cognitive impairment (Hallas, 2011). Due to the limitations in formal care services (Open and Day Care Centers, KAPI/KIFI), the government fosters informal help from families through tax reductions (Daniilidou et al., 2003).

Sweden and Denmark share analogous LTC schemes with both of them offering universal tax-financed coverage to the whole population. The major characteristic of Scandinavian LTC schemes is to ensure that everyone has equal access to care services irrespective of wealth status or place of residence (Lagergren et al., 2016). In Sweden, LTC is regulated at the national level with municipalities being responsible for financing and providing care. Because of the highly developed formal care system, informal care is only perceived as a secondary mean. Measurement of acuity is made through a set of different scales based on ADL and IADL (Fukushima et al., 2010; OECD, 2013). Benefits are paid in-kind by reimbursing LTC expenses and extra cash benefits are offered to family members providing informal care. The Danish LTC system is similar with the exception that the distinction between institutional and home care is less clear than in Sweden (Karlsson et al., 2012).

Central Eastern European countries such as the Czech Republic, Estonia and Slovenia are characterized by a rather low level of LTC services, little regulation by unified laws and high reliance on families. In Czechia, the LTC system is organized through residential, home and community services as well as hospitals for acute care (Österle, 2010). Based on the ADL scale, a cash allowance is paid by the State along four levels ranging from light to very heavy dependence. In specific cases, in-kind benefits are also provided by the health care system (Sowa, 2010). The systems found in Estonia and Slovenia are relatively similar. Benefits are mostly paid in cash and management is left to municipalities (Österle, 2012).

Finally, we separately discuss the case of Austria and Switzerland. The LTC system in Austria is surely one of the most developed within the European Union. It covers the whole population, but supplementary benefits are subject to means test. Social insurance pays for a cash care allowance "Pflegegeld" where the amount depends on the number of hours of care required varying along seven levels. Further, a 24-hours care support benefit is available for those receiving at-home care under means-test conditions. The system is tax-based and managed at the federal and regional levels (Da Roit et al., 2007). Despite the relatively-well developed formal care system, family still plays a central role for shortening the costs. In Switzerland, the LTC system is poorly developed (Fuino and Wagner, 2018b). Mandatory health insurance supports formal care at-home and provides in kind benefits for specific furniture. An old-age care allowance is paid to all 65+ elderly in need of LTC irrespective of their wealth. Especially in institutional care, more than 40% of the costs are left to the households while social security helps those who cannot afford it (Swiss Re, 2014).

In Table 2.1, we summarize for each country the LTC system in place with the type of funding, the organizational levels and the types of functional limitations in ADL and IADL considered for defining dependence levels and benefits. The second column informs about our categorization of the different countries LTC schemes into four groups (family care, State responsibility, subsidiary and none). We describe these classes in the next section (see Conjecture 3).

2.2.2 Research hypotheses

In the following, we derive five conjectures hypothesizing on the characteristics of LTC needs in a sample of European countries. Due to the existing heterogeneity in regulation and in the availability of LTC services, the European continent is a perfect candidate for such study. Taking a cross-country view can provide stronger support to conclusions. The need of LTC is assessed by the measurement of reporting limitations in ADL along the Katz scale. Further, we consider the reporting on the usage of formal LTC services. In our empirical study, we are able to assess these measures throughout selected countries (see Sections 2.3 and 2.4).

Sociodemographic factors The age and the gender are often used in public health research to predict the need of LTC. Germain et al. (2016), find that age and gender are essential drivers

Country	LTC scheme	Funding source	Organizational level	ADL	IADL
Austria	State responsibility	Taxes	Federal and regional	✓	✓
Belgium	Subsidiary	Health insurance	Federal and regional	✓	✓
Czechia	Family care	Health insurance	National, regional and municipal	✓	
Denmark	State responsibility	Taxes	Municipal	✓	✓
Estonia	Family care	Taxes	Regional and municipal	✓	✓
France	Subsidiary	Taxes and private insurance	National	✓	✓
Germany	Subsidiary	Health insurance	Regional and municipal	✓	✓
Greece	Family care	Health insurance	National	✓	✓
Italy	Family care	Taxes	Regional and municipal	✓	
Slovenia	Family care	Taxes	Regional and municipal	✓	
Spain	Family care	Taxes	Regional	✓	
Sweden	State responsibility	Taxes	National and municipal	✓	✓
Switzerland	None	Taxes and health insurance	Federal and regional	✓	

Table 2.1: LTC schemes across selected European countries.

for the risk stratification and for explaining health limitations. Similarly, many authors develop models using age and gender to account for the related heterogeneity in insurance pricing (see, e.g., Pinquet et al., 2011; De Meijer et al., 2011; Fong et al., 2015; Fuino and Wagner, 2018a). Nonetheless, due to limitations of the available data, such studies rarely consider medical factors that are often a major reason for the dependence. For example, the occurrence of diseases like Alzheimer and cancer strongly depends on the individuals' age and gender (see, e.g., Moritz et al., 1995; Letenneur et al., 1999; Smith et al., 2007; Mols et al., 2007). Further, for ages close to life expectancy, the age and the gender are less relevant in shaping LTC needs (Zweifel et al., 2004).

In the United States, spouses are often the first provider of at-home care (Pinquart and Sörensen, 2011). In case of a spouse's absence or inability to provide care, adult children and their spouses represent the second layer for informal care. While receiving help form the partner is usual, help from children is less common and is related to social responsibility of children towards their parents (Zweifel and Struwe, 1996; Courbage and Zweifel, 2011). Societal trends linked to globalization induce increased distances between the childrens' and their parents' locations conducting to higher use of formal care solutions. Freedman and Martin (1998) show that married American elderly have significantly lower prevalence rates in functional limitations than the non-married. On that basis, we consider that elderly living in a two persons household present lower acuity levels than those living in a single person household. Even if it is only a second layer, the presence of children can strongly reduce the reporting of limitations in ADL. For example, Choi (1994) finds that childless elderly are 20.3% to have at least one ADL limitation in comparison to 14.5% for the ones having children not living in the same household and 3.0% for the ones living together with their children. Therefore, we also expect to observe a significant reduction in functional limitations when an elderly has children.

The educational level and the wealth status are often considered as a proxy of the social class of an individual. Freedman and Martin (1999) bring evidence that higher functional limitations come along with lower education levels and might also drive the demand for formal LTC. On a sample of about six thousand American elderlies observed over the period from 1984 to 1993, the authors find a significant reduction in the prevalence of limitations when educational attainments are higher. This conclusion is also supported by Fried et al. (2001) who analyze the effect of many factors including the education level on the frailty of elderly. Their study also shows that wealthier individuals are less likely to present ADL limitations. Many studies link lower financial resources to higher LTC needs (Pollack et al., 2007). Thereby, a partial explanation for such interaction between wealth and health stems from the level of insurance since wealthier elderly often hold supplementary coverage enhancing access to expensive health care services (Mobley et al., 2006). However, these findings are contrasted by other studies indicating no causality of wealth on health (Michaud and van Soest, 2008).

Conjecture 1: *Both (a) the probability to report ADL limitations and (b) the probability to report formal LTC usage are significantly affected by demographic (age, gender, body mass index, daily smoking) and social factors (partner in the household, children, wealth status, education level).*

Pathologies For a long time, mental and physical diseases have been said to strongly impact healthy limitations (Anderson et al., 1993; Guibert and Planchet, 2017) and many studies evidence the effect of dementia. Based on a longitudinal study of 407 elderly without dementia or disability, Lau et al. (2015) find that limitations in functional capacities are highly related to the occurrence of mental diseases. The case of Alzheimer among elderly is well studied since it is the most common cause of dementia. Barberger-Gateau et al. (1992) and Tuokko et al. (2003), discuss the correlation between functional limitations and Alzheimer. Diseases affecting bones and strength always lead to limitations. This assertion probably also holds for physical diseases such as cancer, diabetes, heart failure and high blood pressure. For example, Avis and Deimling (2008) suggest that cancer affects physical functioning. The same is true for elderly diagnosed with diabetes, heart failure and high blood pressure (see, e.g., Kuo et al., 2005; Sinclair et al., 2008; Lesman-Leegte et al., 2009). In the discussion of Conjecture 2, we will argue that cancer might lead to fewer functional limitations than other diseases since, while having higher mortality, it does not always imply dependency (see, e.g., Anderson et al., 1993, where no significant effect of cancer on functional limitations is found).

Conjecture 2: *Medical factors (mental, Parkinson, cancer, musculoskeletal system and other physical diseases) have a strong effect on (a) the probability to report ADL limitations and (b) the probability to report formal LTC usage.*

LTC schemes Our third conjecture directly follows from the discussion in Section 2.2.1. Based on the description of the countries, three clusters of LTC systems emerge. "State responsibility" models are LTC systems in which the government plays a central role. Independently of the wealth status, every citizen in need of care is entitled to receive benefits for financing home- and institution-based care. Mostly funded by taxes, the management and the financing of care are left to local authorities. Municipalities must be aware of the care needs and propose suitable solutions. Such models take inspiration from the Scandinavian health system that offers universal public coverage financed to a large part by taxes (Karlsson et al., 2012). Further, formal community care services (home- and institution-based care) are well developed and cover most of the required LTC. Finally, such models present higher LTC expenditures but satisfy well the population's needs. They also enhance the requirement of care provided by relatives often considered as a secondary solution (Banerjee et al., 2012). The Swedish and the Danish LTC system are part of this category. Further, we also consider the Austrian scheme here although it is a mixture between State responsibility and family care. In fact, "family care" schemes position the family of the dependent elderly as a major actor. Elderly must first request help from their family before relying on State solutions. Further, family members stand as one of the only affordable caregivers inhibiting the development of community services. The government pays limited benefits that are subject to means test and to a minimum acuity level threshold. Such benefits are mostly paid in-cash for facilitating financial retribution to relatives providing care. Family care models are mostly present in Southwestern Europe (Costa-Font, 2012) and to a lesser extent in Central and Eastern Europe (CEE) where most of the population believe that care provided by children is the best option (Carrera et al., 2013). Spain, Italy, Greece as well as CEE countries mostly hold family care systems (Bettio et al., 2006). Finally, "subsidiary" models represent a trade-off between State responsibility and family care models. Such schemes are characterized by well-developed community services and by a strong involvement of the family. Elderly in need of care typically have to announce themselves to the State authority which is responsible to manage the individual health path. However, benefits provided by the system are insufficient to overcome the total LTC costs. To curtail the financial pressure, the family is required to participate in the financing and to provide care (Courbage and Plisson, 2012). Such models mix both formal and informal care. The subsidiary LTC model is found in Germany, France and Belgium (Karlsson et al., 2012). Finally, Switzerland has to be considered separately since no proper scheme LTC system is developed (classified as "none"). Our classification is close to the one proposed by Nies et al. (2013). Table 2.1 summarizes the LTC schemes as laid out above.

Conjecture 3: *Both (a) the probability to report ADL limitations and (b) the probability to report formal LTC usage significantly differ by types of LTC schemes.*

Types of ADL limitations The scale by Katz and Jackson (1963) determines functional limitations among elderly based on six ADL, namely dressing, bathing, getting in and out of bed, toileting, walking and eating (Becker and Reinhard, 2018). Based on medical assessments, LTC dependence follows common patterns. Activities requiring lower extremity strength are affected before upper extremity strength activities (Kempen et al., 1995; Kingston et al., 2012). While a clear distinction is often made among functional limitations, e.g., advanced items such as bathing and dressing and basic items such as toileting and feeding (Rubenstein et al., 1984), only little research is made on the effect of sociodemographic factors on individual ADL limitations (Fong and Feng, 2018). Iwarsson et al. (2009) report results from a survey on adults aged over 75 years carried out in Germany, Hungary, Latvia, Sweden and the UK. Their results reveal that less than 5% report difficulties with feeding while more than 20% report limitations with transferring and bathing. Results from the Assets and Health Dynamics of the Oldest Old Survey (AHEAD) also indicate a hierarchical distinction between ADL. For example, elderly report fewer difficulties with activities such as toileting and eating than with walking and dressing (Himes, 2000).

Conjecture 4: *The probabilities to report limitations along types of ADL are differently affected by demographic, social and medical factors.*

Formal LTC usage Formal LTC corresponds to professional help provided to an elderly in need of assistance with ADL. These services are distinguished between personal care delivered by nurses and domestic tasks provided by non-medical staff (Tennstedt et al., 1993; Comas-Herrera, Adelina, Wittenberg et al., 2006). For example, in the U.S., according to Fox et al. (1999), about 93% elderly require help with transportation while only about 45% need assistance with personal care. Distinguishing between the types of care is important because the lack of caregivers is mostly related to personal care and the required medical staff rather than to domestic tasks (Hussein and Manthorpe, 2005; Fujisawa and Colombo, 2009). Further, elderly can often rely on their partner for domestic tasks, while they require professional help for personal care (Jiménez-Martín and Prieto, 2012).

Conjecture 5: *The probabilities to report formal LTC usage along personal care and domestic tasks are differently affected by demographic, social and medical factors.*

2.3 Dataset and descriptive statistics

To explore Conjectures 1 to 5, we rely on econometric models applied on the records of the Survey of Health, Ageing and Retirement in Europe (SHARE). In Section 2.3.1, we describe the main characteristics of the SHARE dataset and discuss the variables of interest. Then, we present the descriptive statistics and lay out the main figures in Section 2.3.2.

2.3.1 Description of the SHARE dataset

The SHARE dataset provides information on a representative sample of individuals aged 50 years and over living at home (excluding nursing homes and hospitals). Covering twenty-seven European countries, this study is re-conducted every two years and has registered six waves over the period from 2004 to 2018. The records are summarized in modules. Eight modules report original respondents' answers on behavioral risks, children, consumption attitude, demographics, employment, pensions, peak flow, health care and medical assessment. The other three modules contain generated variables for correspondence to international classification standards on physical and mental health, on education and on social network information. The data collection process is managed by local universities that mandate professional polling firms. The process is broken down into two parts: computer-aided personal interviews (CAPI) and paper and pencil medical questions. In the first part, the elderly is assisted by a purpose-trained employee for reporting answers and measuring health indicators such as grip strength and blood pressure. In the second part, the interviewee has to fill in a form treating more personal questions including pathologies.[1] The survey results contain detailed information about respondents' sociodemographic, economic and health characteristics. The individual's *age*, *gender*, *body mass index* (BMI), smoking habits (*daily smoker*) and *country of residence* are typical demographic factors. The survey also contains information on the number of limitations in ADL an elderly has. In our study, we only consider ages from 65 to 99 years, a range where most dependent elderly are found (Balia and Brau, 2014). With respect to the smoking habits and BMI, the records reveal if the respondent has ever been a daily smoker whereas we cluster BMI responses within the six classes defined by the international scale from the World Health Organization (2000), i.e. underweight, normal weight, overweight, moderately obese, severely obese and very severely obese. The country of residence corresponds to one of the twenty-seven European countries within the SHARE dataset. We do not consider behavioral factors such as

[1] The survey methodology highly compares with the one from the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA) reporting information on elderly in the U.S. and England, respectively (Meijer et al., 2008). The method is assessed to provide reliable and comparable records across European countries (Börsch-Supan et al., 2005). This paper uses data from SHARE waves 1, 2, 3 (SHARELIFE), 4, 5 and 6 (Börsch-Supan, 2018a,b,c,d,e,f), see Börsch-Supan et al. (2013) for methodological details.

drinking habits[2] and the loneliness measure since the direction of the causal effect is not clear. For example, medication provided to persons with ADL may prevent alcohol consumption or loneliness may be due to disability.

Regarding the socio-related variables, we first construct a variable *partner in household* for identifying the type of household, i.e. single- and two-persons households. We observe that this factor is highly correlated with the marital status that we do not consider in our sample. Next, we consider the variable of having *children* and we include it as a binary (yes, no) response. Given the substantial difficulties in adding information on the distance to the closest child and the presence of women among children, we exclude them from our study. We introduce the *wealth status* by considering the financial capacity to make ends meet, categorized among four levels, i.e. easily, fairly easily, with some difficulty and with great difficulty. We account for the individuals' *education level* by using the international standard classification of education (ISCED-97) defined by the UNESCO. The SHARE data records the seven levels of the ISCED-97 scale as well as two other classes (persons still studying and others). In our approach, we eliminate individuals with entries in the class "still studying" and "others" representing less than 2% of the data. Further, we cluster the ISCED-97 entries into three groups. The primary education consists in pre-primary and basic education. The secondary education corresponds to the lower, upper and post-secondary education, while the tertiary education contains the first and second stage of tertiary education (United Nations, 1997).

Finally, we consider the physical and mental health status by constructing five medical factors based on the respondents' answers to the medical questions "Doctor told you had: ...". This approach is close to the one used in Balia and Brau (2014) for reporting the effect of diseases on home care utilization among elderly. The *mental diseases* variable is a binary variable that takes the value of one when at least one "yes" appears in the responses to Alzheimer, dementia and emotional disorders. We also consider separately the diagnosis of *Parkinson disease* and *cancer*. Positive responses to hip and femoral fractures, rheumatoid arthritis, osteoarthritis and other fractures are gathered through the *musculoskeletal system diseases* variable. The *other physicals diseases* variable, namely, heart attack, stroke, diabetes, chronic lung disease, cataracts and chronic kidney represent the last cluster.

2.3.2 Descriptive statistics

Our study focuses on the latest wave (wave 6) published in 2018 that contains answers from 39 808 respondents aged between 65 and 99 years. Starting from the raw data, we merge the responses to construct single respondent identifiers. We disregard data from Croatia, Israel, Luxembourg, Poland and Portugal since less than thousand records are available in each of these countries. Due to missing entries in several variables, the inclusion of these countries would lead to the exclusion of many factors of interest. In the end, we remain with 26 331 complete records from the 13 countries discussed in Section 2.2 (also see Table 2.2 below). As part of our data processing,

[2]Many medical studies (e.g., Lang et al., 2007) find a causal link between an excessive alcohol consumption and dependence via neurological diseases and Alzheimer. The effect of alcohol consumption is considered even stronger than the effect of smoking habits and obesity. However, the link between dependence and drinking habits can be reversed: medication due to dependence (or associated diseases) can prohibit alcohol use. Therefore, the current consumption can bias the final result. Further, the past alcohol consumption which could be used is not covered by the SHARE questionnaire. Even though it is possible to restore the past consumption from earlier waves, this approach would significantly reduce the number of observations. Therefore, we do not include alcohol consumption in our analysis.

we are able to fill some missing entries in wave 6 by recovering invariant information provided by respondents in previous waves. Among the respondents, 22 400 are autonomous elderly while 3 931 report at least one limitation in ADL corresponding to a prevalence rate of 14.9%. Across the countries, this percentage of dependent elderly ranges between 9% and 19%.

Demographic, social and medical factors Table 2.2 reports descriptive statistics on the overall, the autonomous ("Aut.") and the dependent ("Dep.") respondents' data (first three columns in each section of the table). When considering the demographic factors, we note that the panel is mostly composed by elderly younger than 85 years (89.4%) with a higher proportion of dependent individuals among the 80+. We observe lower shares of men than women and find the largest disparity (63.3% vs 36.7%) among the dependent elderly. The BMI class with the highest number of records is the "overweight" class. However, we find more "severely obese" and "very severely obese" in the population reporting ADL limitations. About 40% have ever smoked daily. While our records are well distributed among the 13 countries, the family care LTC scheme prevails and represents more than half of the records.

Regarding the social factors, 60% report living with their partner, 90% have children and about 65% are found in the mid-high and high wealth status classes. More than 70% have at least a secondary education background. When considering the dependent population, the main differences arise from the lower share reporting to have their partner in the household (46.6%), from the higher shares in the low (16.2%) and mid-low (30.9%) wealth status classes and from the lower share with tertiary education (12.6%). Finally, when considering medical factors (multiple diseases can be reported), few autonomous individuals report having mental, Parkinson or cancer diseases. Overall, the shares of elderly reporting musculoskeletal system (30.0%) and other physical (38.3%) diseases is significantly higher. We observe higher shares for all diseases when considering the dependent population: more than half announce musculoskeletal system (54.4%) and other physical (63.8%) diseases; 21.1% present mental diseases, 4.5% and 7.5% report Parkinson and cancer diseases, respectively. These findings give first insights on the difference between the autonomous and the dependent populations. In fact, individuals that report limitations in ADL are, in comparison to the autonomous population, older, with a higher BMI and more to live alone. Expectedly, they report more diseases highlighting the importance of considering pathologies when discussing ADL limitations.

Table 2.2 also reports statistics on formal LTC used by elderly presenting functional limitations (last two columns). Among the 3 931 dependent elderly, 2 461 have not reported using any professional services ("w/o FC") while 1 470 have used formal care ("FC"). The distribution of the observations along age classes is noticeably different when comparing dependent elderly using professional services with the others. The share of elderly using formal care increases at ages 80+. For example, they are 22.0% in the segment from 85 to 89 years, to require professional services, which compares to 12.5% of the respondents reporting not using professional help. Further, records on elderly requiring formal help are composed by 70.3% of female and 29.7% of male. This high share of female stems from the higher female life expectancy of dependent elderly (see, e.g., Mathers, 1996). While BMI and smoking habits do not show important differences between both groups, we note strong differences among LTC schemes. Dependent respondents announcing the usage of qualified help are 19.9% living in a State responsibility scheme, 37.8% living in a family care scheme and 38.2% living in a subsidiary scheme. Conversely, the share of elderly not using formal care is much higher in countries with

		Overall	Aut.	Dep.	w/o FC	FC			Overall	Aut.	Dep.	w/o FC	FC
Demographic factors													
Age							**Country of residence**						
65 – 69	%	29.8	32.3	15.8	20.0	8.6	Austria	%	6.8	6.9	6.3	4.5	9.3
70 – 74	%	24.2	25.5	16.7	19.8	11.6	Belgium	%	8.9	8.4	11.5	8.2	17.1
75 – 79	%	20.8	20.7	21.0	22.0	19.3	Czechia	%	9.2	9.1	9.9	12.4	5.8
80 – 84	%	14.6	13.3	21.6	20.5	23.5	Denmark	%	5.5	5.8	3.5	2.7	5.0
85 – 89	%	7.6	6.2	16.0	12.5	22.0	Estonia	%	10.1	9.6	12.8	16.4	6.7
90 – 94	%	2.6	1.8	7.3	4.1	12.7	France	%	7.0	6.7	9.0	7.0	12.3
95 – 99	%	0.4	0.2	1.6	1.1	2.4	Germany	%	6.9	7.0	6.7	5.4	8.8
Gender							Greece	%	7.2	7.4	5.7	6.1	5.0
Male	%	43.0	44.1	36.7	40.9	29.7	Italy	%	8.9	8.9	8.5	9.5	6.8
Female	%	57.0	55.9	63.3	59.1	70.3	Slovenia	%	6.7	6.6	7.5	10.2	3.2
Body mass index							Spain	%	9.3	9.2	9.8	9.5	10.3
Underweight	%	1.3	1.0	2.7	1.9	3.9	Sweden	%	8.2	8.7	5.5	5.4	5.6
Normal weight	%	34.6	35.5	29.6	26.6	34.6	Switzerland	%	5.3	5.7	3.3	2.7	4.1
Overweight	%	42.2	43.2	36.3	37.5	34.2	**LTC scheme**						
Moderately obese	%	16.6	15.9	20.3	21.8	17.9	State responsibility	%	20.5	21.5	15.3	12.6	19.9
Severely obese	%	4.2	3.6	8.0	8.9	6.5	Family care	%	51.4	50.8	54.2	64.0	37.8
Very severely obese	%	1.1	0.8	3.1	3.3	2.9	Subsidiary	%	22.8	22.0	27.2	20.6	38.2
Daily smoker							None	%	5.3	5.7	3.3	2.7	4.1
Yes	%	41.3	42.1	37.0	38.1	35.2							
Social factors													
Partner in household							**Education level**						
Yes	%	60.1	62.4	46.6	55.1	32.4	Primary	%	28.2	26.6	37.7	36.0	40.7
Children							Secondary	%	51.9	52.3	49.7	51.3	46.9
Yes	%	90.3	90.4	89.5	92.3	85.0	Tertiary	%	19.9	21.1	12.6	12.7	12.4
Wealth status													
High	%	35.5	37.2	26.0	23.2	30.7							
Mid-high	%	29.1	29.4	26.9	25.8	28.8							
Mid-low	%	25.4	24.5	30.9	33.4	26.7							
Low	%	10.0	8.9	16.2	17.6	13.8							
Medical factors													
Mental diseases							**Musculoskeletal system diseases**						
Yes	%	8.3	6.0	21.1	18.4	25.6	Yes	%	33.6	30.0	54.4	52.2	58.5
Parkinson disease							**Other physical diseases**						
Yes	%	1.2	0.6	4.5	3.7	5.9	Yes	%	42.1	38.3	63.8	62.5	65.9
Cancer													
Yes	%	5.2	4.8	7.5	7.1	8.0							
Nb. of individuals	N	26 331	22 400	3 931	2 461	1 470		N	26 331	22 400	3 931	2 461	1 470

Note: "Aut." stands for autonomous, "Dep." for dependent, "w/o FC" for without formal care and "FC" for with formal care.

Table 2.2: Descriptive statistics on reported ADL limitations and formal LTC usage by demographic, social and medical factors.

family care policies. The shares are 64.0%, 12.6%, 20.6% and 2.7% for dependent persons not reporting professional help, respectively. Considering social and medical factors, we observe most significant differences with the presence of the partner in the household and diagnosed mental diseases. A share of 55.1% of dependent elderly not using formal care live with their partner while only 32.4% are in two persons households among the professional care takers. Dependent elderly reporting formal care usage also have a higher share of mental diseases (25.6%).

Further, we define the prevalence rate as the ratio of elderly limited in at least one ADL over the total population at the same age. We illustrate prevalence rates separately for male and female respondents in Figure 2.1. Both curves have an exponential shape indicating higher prevalence rates at higher ages. In addition, we observe that female respondents have higher prevalence when compared to male respondents.

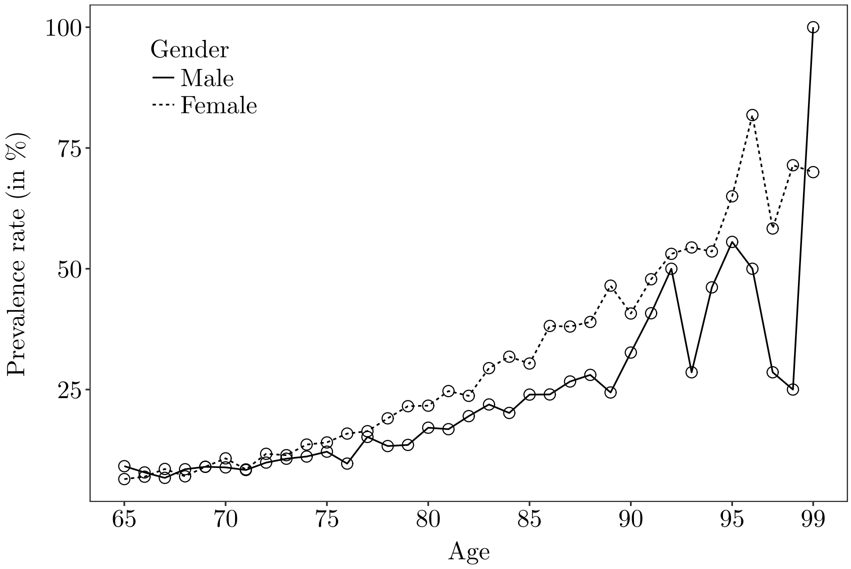

Figure 2.1: Prevalence rates by age and gender.

Types of ADL limitations Among the 3 931 elderly reporting limitations in ADL, we present their distribution along the six ADL in Table 2.3. A single individual can report more than one limitation. We find that the ADL that has been reported the most is dressing (68.3% of the dependent elderly), followed by bathing (54.6%), getting in and out of bed (31.5%), toileting (21.2%), walking (19.0%) and eating (16.5%). This ranking is similar to the one reported by Fox et al. (1999). Although they solely study people diagnosed with dementia coming with more limitations, they find that more than half of their sample is in need of assistance with bathing and dressing while eating, toileting and walking show significantly lower shares.

ADL	N	Share
Dressing	2 686	68.3%
Bathing	2 145	54.6%
Getting in and out of bed	1 238	31.5%
Toileting	833	21.2%
Walking	746	19.0%
Eating	650	16.5%

Note: Shares are based on 3 931 elderly reporting limitations in ADL.

Table 2.3: Descriptive statistics on the reported limitations in specific ADL.

Formal LTC usage In Table 2.4, we detail on the reported formal LTC usage by the 3 931 dependent elderly. Overall, 1 470 (37.4%) report using professional help: 29.0% receive help for domestic tasks and 20.4% get personal care. A lower share uses meals-on-wheels services (9.8%), professional help with other activities (9.3%) and seasonally stays in nursing homes (2.3%). The other 2 461 dependent elderly (62.6%), although reporting limitations in ADL, do not report using any professional services. Recall that the SHARE data only records answers from elderly living at home. This explains the limited usage of nursing home services.

Services	N	Share
With formal care	1 470	37.4%
Domestic tasks	1 139	29.0%
Personal care	803	20.4%
Meals-on-wheels	387	9.8%
Other activities	366	9.3%
Nursing home	92	2.3%
Without formal care	2 461	62.6%

Note: Shares are based on 3 931 elderly reporting limitations with ADL.

Table 2.4: Descriptive statistics on the reported formal care usage of specific services.

2.4 Econometric models and results

In Section 2.4.1, we develop econometric models to study the reporting of limitations in ADL and of formal care usage. Further, we present our results and link them to the research hypotheses in Section 2.4.2. In Section 2.4.3, we discuss the validation of the conjectures and study the robustness.

2.4.1 Model framework

Explanatory variables We are interested in understanding the effect of various factors on the probability to report limitations in ADL and on the probability to report formal LTC. For this purpose, we consider the demographic, social and medical factors introduced in Section 2.3.2. The individual's age (AG) is a numeric variable running from 65 to 99 years. Further, we account for nine binary variables as follows. Four sociodemographic variables indicate whether the respondent is a man or a woman (GE), has ever smoked daily (SM), is living with a partner in the same household (HH) and has children (CH). Five medical factors indicate whether the respondent has mental (MD), Parkinson (PA), cancer (CR), musculoskeletal system (MS) or other physical diseases (PD). The baseline category in these variables is the "no" answer respectively "male" for GE. Finally, we consider four sociodemographic categorical variables giving information on the respondent's body mass index (BM), wealth status (WS), education level (ED) and country of residence's LTC scheme (SC). The range of values taken by the categorical variables are discussed in Section 2.3.1 and summarized in Table 2.5. For each of the variables we define the baseline category and consider "normal weight", "high" wealth status, "primary" education level and "none" for the country's LTC scheme. We include the above variables in our econometric models and use the notation $\boldsymbol{X}$ to refer to the set of variables reported in Table 2.5.

Since our objective is to develop models explaining the effect of the above variables and not to make a prediction, we keep all variables in each model and study their statistical significance. However, to avoid models with too many variables, we only consider relevant interactions that improve our model along the Bayesian information criterion (BIC) values using a backward step-wise selection algorithm. Throughout the various models tested (see below and Table 2.6) the retained interaction terms along BIC are the same. In the models explaining the probability to report ADL limitations, only the gender and age interaction appears relevant. In the probability to report formal LTC models, the interactions of the partner in the household with both age and gender as well as of the children with gender interaction are retained.

Variables	Description	Values
AG	Age	from 65 to 99
GE	Gender	male, female
BM	Body mass index	6 classes from underweight to very severely obese
SM	Daily smoker	yes, no
HH	Partner in household	yes, no
CD	Children	yes, no
WL	Wealth status	high, mid-high, mid-low, low
ED	Education level	primary, secondary, tertiary
MD	Mental diseases	yes, no
PA	Parkinson disease	yes, no
CR	Cancer	yes, no
MS	Musculoskeletal system diseases	yes, no
PD	Other physical diseases	yes, no
SC	LTC scheme	State responsibility, family care, subsidiary, none

Table 2.5: Description and values of the independent variables included in $\boldsymbol{X}$.

Model selection To investigate the probability to report limitations in ADL, we define six binary variables ℓ^j corresponding to the ADL. These variables take the value of one when a limitation in the j^{th} ADL, namely, dressing ($j = 1$), walking ($j = 2$), bathing ($j = 3$), eating ($j = 4$), getting in and out of bed ($j = 5$) and toileting ($j = 6$) is reported. It takes the value of zero otherwise. For studying the reporting of formal care usage, we construct two binary variables f^k identifying the reporting of help for personal care ($k = 1$) and with domestic tasks ($k = 2$). We do not analyze the other formal care services, namely, meals-on-wheels ($k = 3$), other activities ($k = 4$) and nursing ($k = 5$) separately, given their lower prevalence (less than 400 observations, see Table 2.4). With these notations, we can formally write out the dependent variables of interest. The probability to report at least one limitation in any ADL can be written as $\mathbb{P}(\sum_j \ell_i^j > 0)$. The same probability for a specific ADL j is $\mathbb{P}(\ell_i^j = 1)$. The probability to report the usage of at least one formal LTC service is $\mathbb{P}(\sum_k f_i^k > 0)$ and the one related to a specific service k is $\mathbb{P}(f_i^k = 1)$.

A common method for estimating probabilities with econometric models is to use logistic and probit regression models.[3] Denominated as binomial generalized least square models (GLM), these methods differ through the link function used to relate the dependent variable to the independent variables. While the logistic model uses the *logit* function, the *probit* model assumes

[3]Given the fact that SHARE data is longitudinal (i.e., historical snapshots), it is possible to use a survival function to measure the time-to-dependence (e.g., using a Kaplan-Meier estimator or Cox proportional hazards model). Such approach would allow us to validate the consistency with the results obtained by using logistic and probit regressions. However, the observed sample over five waves of SHARE is limited to 2 496 observations, which might yield a high variance of the coefficients. It should also be noted that the different waves have been done at intervals of two years.

a quantile of the standard normal distribution as link function. For comparison, we also consider a linear regression model. To decide which model yields the best fit, we apply four performance measures. They are the log-likelihood, the deviance, the in-sample area under the curve (AUC) and AUC based on 5-fold cross-validation. Better models present higher log-likelihood and AUC but lower deviance. For selecting the best model, we consider both the probability to report at least one limitation in ADL, $\mathbb{P}(\sum_j \ell_i^j > 0)$, and the probability to report the usage of at least one formal LTC service, $\mathbb{P}(\ell_i^j = 1)$. We report the performance measures in both cases in Table 2.6. We find that the probit model performs better than both the logit function and the linear regression model for $\mathbb{P}(\sum_j \ell_i^j > 0)$. We therefore use the probit model when studying the ADL limitations. Using the same statistical procedure, we determine which model fits best to analyze the probability to report a specific ADL limitation $\mathbb{P}(\ell_i^j = 1)$. Here, we also find that the probit model offers the best fit. With regard to $\mathbb{P}(\sum_k f_i^k > 0)$, we find that the logistic model is better suited (see Table 2.6). When considering the probability to report formal LTC with personal care or with domestic tasks, the logistic regression model also outperforms the other models.

	$\mathbb{P}(\sum_j \ell_i^j > 0)$			$\mathbb{P}(\sum_k f_i^k > 0)$		
	probit	logit	linear	probit	logit	linear
In-sample						
Log-likelihood	−9 156.91	−9 167.86		−2 189.99	−2 185.57	
Deviance	18 313.83	18 335.72		4 379.98	4 371.14	
AUC	0.7883	0.7880	0.7869	0.7600	0.7603	0.7590
5-fold cross-validation						
AUC	0.7872	0.7867	0.7854	0.8027	0.8027	0.7614

Table 2.6: Performance measures for probit and logit link functions as well as linear regression models on the probability to report limitations in ADL and usage of formal LTC.

Specification of the models Following on the above discussion, we introduce two models to study the probability to report ADL limitations. In model (1), we study the probability to report at least one limitation with ADL which can be written as

$$\text{probit}\left[\mathbb{P}\left(\sum\nolimits_j \ell_i^j > 0\right)\right] = \alpha + \boldsymbol{\beta}\boldsymbol{X}_i + \beta_{AG\cdot GE} AG_i \cdot GE_i, \tag{1}$$

where α is the intercept, $\boldsymbol{\beta}$ are the vectors of regression coefficients for the variables $\boldsymbol{X}_i$ (see Table 2.5) and $\beta_{AG\cdot GE}$ is the regression coefficient for the age-gender interaction term for a respondent i. In model (2), we detail the first model by identifying the relation between the covariates and specific functional limitations j. The considered regression equations for $j = 1, \dots, 6$ are:

$$\text{probit}\left[\mathbb{P}\left(\ell_i^j = 1\right)\right] = \alpha + \boldsymbol{\beta}\boldsymbol{X}_i + \beta_{AG\cdot GE} AG_i \cdot GE_i. \tag{2}$$

We evaluate this second model separately for each type of ADL, i.e. $j = 1, \dots, 6$.

Similarly, we develop two models for studying the probability to report formal LTC usage.

Through model (3), we analyze the probability of using at least one formal LTC service:

$$\text{logit}\left[\mathbb{P}\left(\sum_k f_i^k > 0\right)\right] = \alpha + \boldsymbol{\beta X_i} + \beta_{HH\cdot AG} HH_i \cdot AG_i + \beta_{HH\cdot GE} HH_i \cdot GE_i + \beta_{CD\cdot GE} CD_i \cdot GE_i. \tag{3}$$

Again, α represents the intercept and $\boldsymbol{\beta}$ the regression coefficients for the covariates $\boldsymbol{X}_i$. The coefficients $\beta_{HH\cdot AG}$, $\beta_{HH\cdot GE}$ and $\beta_{CD\cdot GE}$ are related to the three interaction terms considered. We refine model (3) by considering separate types of formal LTC in model (4), i.e. for $k = 1, 2$:

$$\text{logit}\left[\mathbb{P}\left(f_i^k = 1\right)\right] = \alpha + \boldsymbol{\beta X_i} + \beta_{HH\cdot AG} HH_i \cdot AG_i + \beta_{HH\cdot GE} HH_i \cdot GE_i + \beta_{CD\cdot GE} CD_i \cdot GE_i. \tag{4}$$

This model is evaluated twice, separately for the probabilities to report use of personal care ($k = 1$) and help with domestic tasks ($k = 2$).

2.4.2 Results and discussion

In this section, we report the obtained results when applying models (1) to (4) on the data described in Section 2.3. We discuss our findings along the Conjectures 1 to 5 introduced in Section 2.2.2. The numerical outcomes are presented in Tables 2.7 and 2.8. In each table, we report the coefficient estimates with their standard error in parentheses. We assess the statistical significance through p-values using the notations "." for a p-value below 0.1, "*" for a p-value below 0.05, "**" for a p-value below 0.01 and "***" for a p-value lower than 0.001. The study of ADL limitations in models (1) and (2) is based on the full set of 26 331 observations. The usage of LTC services in models (3) and (4) is analyzed on the 3 931 records related to dependent elderly (cf. Table 2.2).

Probability to report limitations in ADL

Following the results presented in Table 2.7, we discuss our results along the conjectures.

Conjecture 1a When considering the results obtained for model (1) on the probability to report at least one limitation in ADL, we find that both age and gender are important. The coefficient for the age effect is positive and statistically significant ($\beta_{AG} = 0.033$) indicating that the probability to report limitations in ADL increases with the age. This probability appears, at first glance, to be lower for female than for male ($\beta_{GE} = -1.402$). However, when adding the age and gender interaction, we note that men indeed have higher probability than female to report functional limitations for ages below 78 years. Conversely, at higher ages, women present a higher probability. This finding is in line with the conclusions by Strauss et al. (2003) who find that women have more limitations in ADL than men at ages above 77 years. Focusing on the BMI, we observe that the probability to be dependent significantly increases when deviating from normal weight. Most extreme results are found for very severely obese persons ($\beta_{BM} = 0.983$). Having ever smoked daily does not influence the probability. Further, we find that the presence of the partner in the household significantly decreases the probability to report at least one limitation ($\beta_{HH} = -0.084$). In fact, spouses are often the first providers of care and their presence in the household surely leads to an underestimation of personal frailty (Pinquart and Sörensen, 2011). Indeed, partners actively participate in domestic tasks and provide help with other activities. Our result also relates to the finding in Nielsen et al. (1972) stating that elderly living together with their partner are less likely to require help in an institution. Next, we remark that the sole information of having children does not impact

Model	Model (1)	Model (2)					
	Dependent	Dressing	Walking	Bathing	Eating	In/out of bed	Toileting
Intercept	−4.299 (.195) ***	−3.920 (.210) ***	−4.890 (.348) ***	−5.569 (.253) ***	−4.657 (.351) ***	−4.664 (.298) ***	−4.740 (.334) ***
Age	0.033 (.002) ***	0.025 (.003) ***	0.029 (.004) ***	0.044 (.003) ***	0.028 (.004) ***	0.027 (.003) ***	0.028 (.004) ***
Gender (baseline: Male)							
Female	−1.402 (.226) ***	−1.411 (.246) ***	−1.343 (.385) ***	−1.299 (.286) ***	−1.451 (.405) ***	−0.694 (.327) *	−1.171 (.374) **
Gender × Age	0.018 (0.003) ***	0.017 (.003) ***	0.017 (.005) ***	0.018 (.004) ***	0.018 (.005) ***	0.010 (.004) *	0.016 (.005) **
Body mass index (baseline: Normal weight)							
Underweight	0.437 (.082) ***	0.469 (.108) ***	0.458 (.089) ***	0.440 (.110) ***	0.402 (.102) ***	0.357 (.111) **	0.357 (.111) **
Overweight	0.043 (.025) .	0.060 (.028) *	−0.059 (.042)	−0.053 (.031) .	−0.102 (.044) *	0.041 (.036)	−0.016 (.041)
Moderately obese	0.266 (.031) ***	0.323 (.034) ***	0.039 (.053)	0.104 (.038) **	−0.039 (.056)	0.198 (.044) ***	0.098 (.051) .
Severely obese	0.573 (.047) ***	0.625 (.050) ***	0.254 (.077) **	0.302 (.058) ***	0.020 (.091)	0.354 (.065) ***	0.328 (.074) ***
Very severely obese	0.983 (.080) ***	1.136 (.081) ***	0.735 (.110) ***	0.799 (.090) ***	0.450 (.130) ***	0.757 (.098) ***	0.628 (.114) ***
Daily smoker (baseline: No)							
Yes	0.041 (0.023) .	0.034 (.025)	0.028 (.040)	0.081 (.029) **	−0.001 (.042)	−0.016 (.034)	0.051 (.039)
Partner in household (baseline: No)							
Yes	−0.084 (.024) ***	−0.021 (.026)	−0.042 (.040)	−0.138 (.029) ***	−0.056 (.042)	−0.032 (.033)	−0.027 (.039)
Children (baseline: No)							
Yes	0.019 (.035)	0.030 (.039)	−0.112 (.055) *	−0.012 (.042)	0.039 (.062)	−0.037 (.049)	−0.107 (.054) *
Wealth status (baseline: High)							
Mid-high	0.091 (.028) **	0.083 (.031) **	0.047 (.049)	0.079 (.035) *	0.065 (.051)	0.085 (.042) *	0.074 (.048)
Mid-low	0.235 (.029) ***	0.227 (.032) ***	0.129 (.050) **	0.209 (.036) ***	0.148 (.053) **	0.201 (.042) ***	0.154 (.049) **
Low	0.385 (.038) ***	0.379 (.041) ***	0.309 (.060) ***	0.341 (.045) ***	0.333 (.064) ***	0.346 (.051) ***	0.325 (.058) ***
Education level (baseline: Primary)							
Secondary	0.013 (.024)	0.027 (.027)	−0.033 (.039)	−0.019 (.029)	0.038 (.042)	0.032 (.033)	−0.002 (.038)
Tertiary	−0.139 (.034) ***	−0.107 (.037) **	−0.232 (.061) ***	−0.234 (.044) ***	−0.170 (.064) **	−0.104 (.050) *	−0.140 (.058) *
Mental diseases (baseline: No)							
Yes	0.619 (.032) ***	0.604 (.033) ***	0.510 (.045) ***	0.692 (.034) ***	0.675 (.045) ***	0.609 (.038) ***	0.673 (.042) ***
Parkinson disease (baseline: No)							
Yes	1.046 (.076) ***	0.952 (.076) ***	0.639 (.095) ***	0.847 (.080) ***	0.898 (.089) ***	0.939 (.082) ***	0.835 (.088) ***
Cancer (baseline: No)							
Yes	0.219 (.043) ***	0.162 (.047) ***	0.203 (.067) **	0.275 (.050) ***	0.248 (.069) ***	0.295 (.056) ***	0.257 (.064) ***
Musculoskeletal system diseases (baseline: No)							
Yes	0.400 (.022) ***	0.383 (.024) ***	0.196 (.037) ***	0.243 (.027) ***	0.104 (.039) **	0.248 (.031) ***	0.205 (.036) ***
Other physical diseases (baseline: No)							
Yes	0.356 (.021) ***	0.326 (.024) ***	0.277 (.037) ***	0.378 (.027) ***	0.316 (.039) ***	0.315 (.031) ***	0.281 (.036) ***
LTC scheme (baseline: None)							
State responsibility	0.080 (.057)	0.096 (.063)	0.446 (.136) **	0.091 (.078)	0.191 (.110) .	0.222 (.103) *	0.242 (.111) *
Family care	0.058 (.055)	0.011 (.060)	0.499 (.132) ***	0.215 (.074) **	0.103 (.107)	0.423 (.098) ***	0.287 (.107) **
Subsidiary	0.256 (.055) ***	0.193 (.061) **	0.418 (.134) **	0.343 (.075) ***	0.149 (.109)	0.256 (.101) *	0.159 (.110)
N	26 331	26 331	26 331	26 331	26 331	26 331	26 331

Notes: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.7: Results for regression models (1) and (2) on the probability to report limitations in ADL.

the probability of being limited in ADL. Even if a non-significant effect of children is also found by Van Houtven et al. (2015), we mitigate such assertion since we believe that more precise data like, e.g., the distance between parents and children needs to be considered (Courbage and Costa-Font, 2015). Our outcomes reveal that wealthier elderly are more likely to be autonomous. This is particularly visible in elderly from the high wealth class (baseline) when compared to the low-income class ($\beta_{WL} = 0.385$). Following on that, we also note that individuals with a higher education level report fewer difficulties with ADL (Freedman and Martin, 1999; Fried et al., 2001; Bonsang, 2009; Fonseca and Zheng, 2011; Di Gessa and Grundy, 2014). As an answer to Conjecture 1a, we conclude that the age, the gender, the BMI, the presence of the partner in the household, the wealth status and the education level significantly affect the probability to report at least one limitation in ADL.

Conjecture 2a The medical information about the diagnosis of mental, Parkinson, cancer, musculoskeletal and other physical diseases are significantly related to the reporting of functional limitations. Among the considered diseases, Parkinson has the strongest impact ($\beta_{PA} = 1.046$), reducing the motor system of the human body and yielding four main symptoms: tremor (shaking), rigidity (stiffness), bradykinesia (slowness of movements) and postural instability (National Institute of Health, 2014). These symptoms directly induce disabilities of performing ADL and entail dependence. Mental diseases have the second highest effect on the occurrence of limitations ($\beta_{MD} = 0.619$). As discussed in Section 2.2.2, Alzheimer is among the most diagnosed mental diseases and leads to strong functional impairments (Barberger-Gateau et al., 1992; Tuokko et al., 2003). According to the World Health Organization (2017), mental disorders are a major contributor to the global burden of diseases and in particular among elderly. Further, the presence of musculoskeletal system and other physical diseases increases the probability since they directly relate to functional limitations ($\beta_{MS} = 0.400$ and $\beta_{PD} = 0.356$). Intuitively, one would expect that physical diseases have a stronger effect on ADL limitations than mental diseases. However, two main components of the factor "other physical diseases" are stroke and heart attack, which rather cause death than long-term disability in aged people. Finally, the effect of cancer is found to be the smallest ($\beta_{CR} = 0.219$). Similarly to strokes and heart attacks, cancer more rapidly leads to death when compared to other diseases, sometimes avoiding elderly to experience dependence (see, e.g., Anderson et al., 1993; Fonseca and Zheng, 2011).

Conjecture 3a In our analysis of the LTC scheme's effect on the probability to report limitations in ADL, we take Switzerland ("none") as the baseline for comparison. Given that Switzerland has not developed any particular scheme, it is the perfect reference candidate (Swiss Re, 2014). From the numerical results, we remark that elderly living in countries with subsidiary schemes, i.e. Belgium, France and Germany, show significantly higher probability to report limitations in ADL. Such result may be mostly due to Belgium and France gathering a higher share of dependent persons (11.5% and 9.0%, see Table 2.2 in Section 2.3.2). For both State responsibility and family care schemes, we do not observe any significant difference in the probability of reporting functional limitations when compared to Switzerland. This highlights a quite similar pattern and rather shows that the reporting of ADL limitations is not influenced by the LTC scheme. In fact, this could be expected since the population health should not relate to a specific LTC policy.

Conjecture 4 Our fourth conjecture investigates on potential differences existing along ADL types. For all the types of ADL, the probability to report limitations increases at higher ages, for females and for a lower wealth status as well as for elderlies with mental, Parkinson, cancer, musculoskeletal system and other physical diseases. However, we note differences in the other factors. For example, persons defined as moderately obese have difficulties with dressing, bathing and getting in and out of bed while that characteristic is not significant in walking, eating and toileting. Severely obese persons do not present a significant limitation with eating while very severely obese significantly infer assistance with all activities. Detailing limitations by type of ADL also reveals that the presence of the partner in the household reduces the probability to report limitations with bathing while it has no significant effect on the other activities. Another interesting result comes from the education level. Only persons having a tertiary level of education show a lower probability to report difficulties with walking and bathing activities ($\beta_{ED} = -0.232$ and $\beta_{ED} = -0.234$) when compared to the primary education baseline. Such result may come from the education level being related to the type of profession done during the working life. Indeed, there is strong evidence that blue-collar workers present higher functional limitations at high ages than white-collar workers (see, e.g., Dong et al., 2011). Finally, the LTC schemes affect only heterogeneously the functional limitations impeding a clear interpretation.

Probability to report usage of formal LTC

Below we comment on the results displayed in Table 2.8 along the Conjectures 1b, 2b, 3b and 5 related to the usage of formal care services.

Conjecture 1b The results from model (3) show that the probability to report formal LTC usage is positively driven by the individual's age. Due to the increase in ADL limitations, older individuals have higher chances to request formal LTC ($\beta_{AG} = 0.062$). Further, because of their lower mortality, females present higher probability to use formal LTC than males ($\beta_{GE} = 0.766$). Focusing on the other sociodemographic factors such as the BMI, the smoking habits, the presence of children, the wealth status and the education level, we observe that they do not significantly affect the probability. Nevertheless, we would have expected a positive effect along higher wealth status or higher education level since these segments might ask for more professional services. In fact, while fewer functional limitations are reported by higher socio-economic classes, in case of dependence, they may hold financial means or supplementary insurance coverage yielding easier access to formal LTC (Mobley et al., 2006). An interesting result comes from the analysis of the household composition. Persons living with their partner in the household present a significantly lower probability to report formal LTC ($\beta_{HH} = -2.754$). Further, the men's probability is significantly lower when having their spouse in the household than for women (positive coefficient $\beta_{HH \cdot GE} = 0.700$). This confirms that spouses and mostly women are often the first provider of at-home care (Pinquart and Sörensen, 2011).

Conjecture 2b Recall that when analyzing the probability to report functional limitations in the previous section, we have found that diseases yield higher limitations. We are now interested in understanding how diseases can affect the use of formal LTC. Focusing on the values of the different diseases' coefficients, we find the highest impact and statistical significance for mental diseases ($\beta_{MD} = 0.364$) and Parkison ($\beta_{PA} = 0.517$). We note a clear distinction between limitations in ADL and usage of formal LTC. While these diseases highly influence the number

Model	Model (3)			Model (4)					
	Formal LTC			Personal care			Domestic tasks		
Intercept	−4.760	(.625)	***	−6.214	(.723)	***	−5.330	(.659)	***
Age	0.062	(.007)	***	0.062	(.008)	***	0.059	(.007)	***
Gender (baseline: Male)									
Female	0.766	(.242)	**	0.696	(.274)	*	0.783	(.255)	**
Body mass index (baseline: Normal weight)									
Underweight	0.287	(.231)		0.503	(.233)	*	0.389	(.234)	.
Overweight	−0.099	(.091)		0.005	(.104)		−0.135	(.098)	
Moderately obese	−0.073	(.109)		−0.101	(.128)		−0.048	(.118)	
Severely obese	−0.182	(.153)		−0.039	(.184)		0.042	(.162)	
Very severely obese	0.249	(.223)		0.498	(.258)	.	0.317	(.239)	
Daily smoker (baseline: No)									
Yes	0.127	(.086)		−0.048	(.099)		0.157	(.093)	.
Partner in household (baseline: No)									
Yes	−2.754	(1.026)	**	−1.369	(1.011)		−3.907	(.957)	***
Partner in household × Age	0.021	(.011)	*	0.011	(.012)		0.030	(.012)	*
Partner in household × Gender	0.700	(.171)	***	0.011	(.200)		1.197	(.189)	***
Children (baseline: No)									
Yes	0.020	(.208)		0.208	(.247)		0.151	(.230)	
Children × Gender	−0.847	(.256)	***	−0.663	(.291)	*	−0.821	(.274)	**
Wealth status (baseline: High)									
Mid-high	0.027	(.102)		0.094	(.114)		0.079	(.108)	
Mid-low	−0.109	(.104)		0.005	(.120)		−0.117	(.112)	
Low	0.041	(.128)		0.049	(.151)		0.111	(.138)	
Education level (baseline: Primary)									
Secondary	−0.103	(.083)		−0.214	(.094)	*	−0.047	(.089)	
Tertiary	−0.067	(.125)		−0.301	(.146)	*	0.071	(.134)	
Mental diseases (baseline: No)									
Yes	0.364	(.090)	***	0.502	(.098)	***	0.300	(.096)	**
Parkinson disease (baseline: No)									
Yes	0.517	(.174)	**	0.515	(.183)	**	0.687	(.181)	***
Cancer (baseline: No)									
Yes	0.271	(.139)	.	0.153	(.160)		0.066	(.154)	
Musculoskeletal system diseases (baseline: No)									
Yes	0.077	(.077)		−0.141	(.088)		0.224	(.083)	**
Other physical diseases (baseline: No)									
Yes	0.212	(.079)	**	0.205	(.091)	*	0.237	(.085)	**
LTC scheme (baseline: None)									
State responsibility	−0.011	(.216)		0.439	(.243)	.	0.106	(.227)	
Family care	−1.206	(.209)	***	−0.803	(.240)	***	−1.210	(.222)	***
Subsidiary	0.190	(.209)		0.236	(.237)		0.356	(.219)	
N	3 931			3 931			3 931		

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.8: Results for regression models (3) and (4) on the probability to report usage of formal LTC.

of functional limitations (cf. Table 2.7), they do not necessarily entail more formal LTC usage. Indeed, formal care services are often requested in case of diseases that importantly affect functional limitations. Caring for patients with dementia or Parkinson requires complementing informal with formal care especially due to behavioral disturbances (Zarit, 1996; Jiménez-Martín and Prieto, 2012). Further, we highlight that, in comparison to the above diseases, less formal care is sought in the case of cancer, musculoskeletal system and other physical diseases ($\beta_{CR} = 0.271$, $\beta_{MS} = 0.077$ and $\beta_{PD} = 0.212$) since those patients can often be better handled by the informal caregivers without involving an unbearable burden in all cases.

Conjecture 3b Formal LTC usage can be affected by the type of LTC policy. Our results show that living in family care scheme countries differently affects the probability to use formal LTC than living in places where State responsibility or subsidiary schemes are in place. In fact, we find that elderly living under family LTC schemes have significantly lower probabilities to report formal care than the others ($\beta_{SC} = -1.206$). This outcome becomes even more interesting when we compare it to the results observed in the functional limitations where such scheme has no particular effect on reducing the probability to report limitations (cf. Table 2.7). Therefore, we can draw various conclusions. First, policies surely affect the delivery of formal care. In family care countries, the State is often explicitly relying on help provided by family members since the development of private services LTC is low (see, e.g., Bolin et al., 2008; Saraceno, 2010 and our discussion in Section 2.2.1). For example, in Italy most of the allowances are paid only to severely dependent elderly forcing families to rely first on informal care if they cannot afford professional care expenses. Second, culture plays an important role (Gentili et al., 2017). While relying on professional care appears to be common in countries with State responsibility and subsidiary schemes, using formal care services is often avoided in family care systems. Third, the development of specialized infrastructures, their availability, as well as the training of at-home professional caregivers is a condition to the usage of formal LTC. Scarce availability is often an issue in family care systems when formal LTC is less developed.

Conjecture 5 Our last conjecture is interested in the factors' differences between the two types of formal LTC, personal care and domestic tasks. In fact, personal care is often both psychologically and mentally more demanding for the informal caregiver than domestic tasks that relate to common household responsibilities. While our results do not show particular differences when comparing the effect of the age, the gender, the BMI, the smoking habits and the wealth status, we observe diverging results in the other factors. Having the partner in the household highly reduces the request of professional services for domestic tasks ($\beta_{HH} = -3.907$). However, this reduction effect is hindered when the elderly requiring LTC is a woman ($\beta_{HH \cdot GE} = 1.197$). Similar outcomes are found in the literature on informal LTC. For example, Navaie-Waliser et al. (2002) find a higher share of women providing help with ADL than men (e.g., 30% of women with a dependent partner are providing help with bathing, while this ratio decreases to 13% for men). Regarding the education levels, we note that the probability to use personal care is reduced in elderly with a higher education level ($\beta_{ED} = -0.214$ for secondary and $\beta_{ED} = -0.301$ for tertiary education levels, respectively). This effect is not significant for domestic tasks. Looking at the pathologies, we observe that mental diseases and Parkinson increase the probability to request professional help with both personal care and domestic tasks. Being diagnosed with cancer does not influence these probabilities. Further, major differences appear when considering musculoskeletal system diseases and other physical diseases which significantly affect the probability to use help with domestic tasks. Finally, we find that the probability to request formal LTC with personal care and domestic tasks is significantly lower in family care schemes. This effect is stronger for domestic tasks ($\beta_{SC} = -1.210$) than for personal care ($\beta_{SC} = -0.803$). In fact, the offer of services in domestic tasks is broader when compared to personal care since fewer qualifications are needed.

2.4.3 Validation of conjectures and discussion

Using the results from Section 2.4.2 based on wave 6 data from SHARE, we assess whether the conjectures defined in Section 2.2.2 hold. Then, we extend our validation procedure by comparing results with historical data. In fact, our analysis has been performed so far on the most recent wave 6. To evaluate the robustness of our conclusions, we apply the regression models (1) and (3) as far as possible on previous waves of the SHARE data. This allows us to present results (see also the Appendix) validating or refuting conjectures on the basis of earlier waves. Table 2.9 presents a summary of our final conclusions on the conjectures with cross-wave comparisons.

Assessment with wave 6 data From the findings in Sections 2.4.2 and 2.4.2 our conclusions are as follows: the first conjecture questions the effect of sociodemographic factors on (1a) the probability to report ADL limitations and on (1b) the probability to report formal LTC. For the probability to report ADL limitations, we find that, on the exception of smoking habits and the presence of children, all factors have a significant effect, therefore confirming Conjecture 1a. For the probability to report formal LTC, the answer is mitigated because BMI, smoking habits, wealth status and education level appear to have no significant impact. For this reason, we only partially validate the effect of sociodemographic factors in Conjecture 1b. Our second conjecture wonders on the effect of pathologies. We conclude that an individual's pathology strongly affects functional limitations, thus confirming Conjecture 2a. With less emphasis (not all medical factors have a highly significant effect) our results also yield the validation of Conjecture 2b for pathologies affecting the usage of formal LTC. Regarding Conjectures 3a and 3b, our outcomes yield for partial agreement. For the reporting

	Wave 6	Wave 5	Wave 4	Wave 2	Wave 1
Probability to report ADL limitations					
Conjecture 1a	✓	✓	✓	✓	✓
Conjecture 2a	✓	✓	✓	✓	✓
Conjecture 3a	(✓)	✗	✗	✗	✗
Conjecture 4	✗				
N	26 331	17 031	11 975	7 274	6 269
Probability to report formal LTC usage					
Conjecture 1b	(✓)	(✓)		(✓)	(✓)
Conjecture 2b	(✓)	✓		✓	✗
Conjecture 3b	✓	✗		✗	✗
Conjecture 5	✓				
N	3 931	2 542	1 960	1 092	983

Note: "✓" proven, "(✓)" partially proven "✗" refuted.

Table 2.9: Summary results on the validation of the conjectures through waves 1, 2, 4, 5 and 6.

of functional limitations, subsidiary LTC schemes are the only to produce a difference when compared to countries with no LTC scheme. These results may however be an artefact to the higher prevalence of dependent persons in those countries in our data. We consider partial validation of (3a). For the probability to use formal LTC, elderly living in countries with family care schemes show a significantly and strongly lower probability when compared to the baseline, while living in one of the two other LTC scheme countries has no impact. The fourth conjecture questions the importance of analyzing ADL separately. Our study does not reveal important differences in the factors affecting the probability to report any of the ADL individually. Our main comments concern the difference in the effect of the BMI depending on the activity and on the presence of the partner in the household that essentially reduces the probability for bathing limitations. Finally, Conjecture 5 considers the distinction of formal care between personal care and domestic tasks. We find clear evidence on the role of the partner in the household reducing the probability to require help with domestic tasks while no significant effect appears when considering personal care. The education level is significantly reducing the probability to report professional help usage for personal care while it does not affect the one for domestic tasks. Further, living in a country with a family care policy has a higher reduction effect on the probability to report help for domestic tasks than for personal tasks. From these observations, we conclude on the importance to distinguish among types of formal LTC and validate Conjecture 5. We summarize the above conclusions in the column entitled “Wave 6” in Table 2.9.

Assessment with historical data To further confirm our hypotheses and evaluate the robustness of our results, we use data from previous SHARE studies where waves 1, 2, 4 and 5 (years 2004, 2006, 2012, 2015) provide some data on elderly reporting limitations in ADL and usage of professional services. For each wave, we construct measures that are the closest and most comparable with wave 6. In fact, the SHARE structure has developed along the years including more countries and pathologies as well as asking more precise questions on the requirements of formal LTC. For example, the countries Czech Republic, Estonia, Greece and Slovenia included in wave 6 are absent in previous waves. Further, information on mental diseases as well as on rheumatoid arthritis, osteoarthritis, chronic kidney disease and other fractures are not reported in earlier waves. Finally, we notice small differences several questions related to receiving help from professional caregivers. For example, in waves 5 and 6 the question on receiving professional help with domestic tasks specifies “in your own house”, while such specification does not appear in waves 1 and 2. No information on these questions is reported in wave 4. We will not be able to test Conjectures 4 and 5 on historical data since the various regressions include too many different definitions among waves making any results hardly comparable. The detailed results when applying models (1) and (3) on the data of the waves 1, 2, 4 and 5 respectively 1, 2 and 5 are reported in Tables 2.10 and 2.11 of the Appendix.

When considering the probability to report ADL limitations, we can validate Conjecture 1a in previous waves. However, it must be noted that the gender coefficient is not significant when the age-gender interaction is included. Such result is unexpected since most research finds significantly higher limitations for men than women at high ages (see our discussion in Section 2.2.2 and the findings in Section 2.4.2). Nevertheless, the age-gender interaction is significant and, when removing that interaction from the model, the gender factor appears to be positive and highly significant as well. While we validate Conjecture 2a among all waves, the LTC scheme has no particular effect on the probability to report ADL limitations yielding rejection of the Conjecture 3a. When considering the probability to report formal LTC usage, we merely find a non-significant effect of the gender leading to partial validation of Conjecture 1b. No conclusion can be drawn on data from wave 4 since formal care has not been recorded that year. Further, we disprove Conjecture 2b in wave 1 (no pathology has a statistically significant effect) and validate it for waves 2 and 5. Note at this point that wave 1 presents less than 1 000 observations on dependent elderly. Finally, for all historical waves, we reject Conjecture 3b since we do not find any effect of the LTC scheme on the probability to report formal care.

The above conclusions are summarized in Table 2.9. For comparison and as confidence indicator the underlying number N of records is indicated for each wave only a lower number of observations is available in earlier waves. Not all factors have the same significance over the 14-years period between waves 1 and 6, but we find similar results along our conjectures with some difficulties to obtain significant results in earlier waves. Although, when challenging our results among waves is subject to bias due to differences in the recorded variables, the analysis performed in this section underlines several findings but also highlights the need of cautiousness in deriving strict conclusions.

2.5 Conclusion

In this paper, we assess the probability to report ADL limitations and the probability to use formal LTC among elderly aged 65+ years in Europe. Our study is based on 26 331 individuals from which 3 931 report functional limitations and 1 470 use professional care services. The data comes from the sixth wave of the SHARE dataset encompassing 13 European countries. We address five conjectures hypothesizing on the importance of demographic, social, medical and policy factors as well as of the ADL and types of formal care in the determination of the aforementioned probabilities.

From a probit regression model, we find that the age, the gender, the BMI, the presence of the partner in the household, the wealth status, the education level and the pathology highly influence the reporting of ADL limitations. First, women aged above 80 years report more limitations than men. Second, while we observe no effect of having children on the reported functional limitations, this result is reversed when a partner lives in the household. Third, persons with higher wealth and higher education level report lower difficulties with ADL. Thereafter, our findings show that pathologies importantly increase functional limitations. However, the effect is lower for people diagnosed with cancer when compared to mental diseases or Parkinson. Finally, the LTC policy of the country of residence has little influence on the reporting.

Based on a logit regression model, we study the covariates influencing the probability to report formal LTC usage. In comparison to functional limitations that can be observed, the usage of professional care is subject to the availability of informal care. In our results, we find that older women appear to require more formal care than men. While some socio-related factors such as the BMI, the smoking habits and the wealth status have no influence, we observe a lower probability for men when there is a partner in the household. This observation highlights the importance of the spouse in providing informal LTC and limiting the demand of professional care. From the pathologies, we find that mental diseases and Parkinson have the highest effect on using formal LTC. In comparison to other diseases, they often require qualified caregivers. Our results also highlight that a cancer diagnosis does not entail significantly more formal care. Further, we observe that elderly living in family care schemes report significantly less formal care usage than those from other cultural backgrounds or countries with different policy arrangements. Finally, by distinguishing formal care along personal and domestic tasks, we find that the partner in the household mostly helps in reducing the probability to demand professional help with domestic tasks.

The SHARE dataset has enabled us to identify functional limitations in ADL and formal LTC usage in a comprehensive cross-country study. While offering an extensive analysis along demographic, social, medical and policy factors, various streams of research could extend the present study. First, the delivery of informal care by relatives surely needs more investigation. For example, providing informal care can also be linked to moral hazard issues in particular linked to the expected inheritance from children (see also Courbage and Zweifel, 2011). Second, in our study we classified pathologies in five main groups and consider only the main pathology of the elderly, i.e. we do not consider co-morbidity. Knowing that over 12 000 diagnoses are listed in the ICD classification, we are convinced that, with the appropriate data, a more detailed analysis on the importance of pathologies in LTC could be performed (see also Price et al., 2005). Finally, our findings can foster the development of LTC insurance by giving insights for the underwriting standards used in future products and in different LTC policy environments across Europe.

Bibliography

Aguilova, L., H. Sauzéon, É. Balland, C. Consel, and B. N'Kaoua, 2014, Grille AGGIR et Aide à la Spécification des Besoins des Personnes Agées en Perte d'Autonomie, *Revue Neurologique*, 170(3):216–221.

Anderson, J. J., J. M. Anthony, Y. Zhang, P. W. F. Wilson, M. Kelly-hayes, P. A. Wolf, B. E. Kreger, and W. B. Kannel, 1993, The Effects of Specific Medical Conditions on the Functional Limitations of Elders in the Framingham Study, *American Public Health Association*, 84(3):351–358.

Avis, N. E. and G. T. Deimling, 2008, Cancer Survivorship and Aging, *Cancer*, 113(12):3519–3529.

Balia, S. and R. Brau, 2014, A Country for Old Men? Long-Term Home Care Utilization in Europe, *Health Economics*, 23(10):1185–1212.

Banerjee, A., T. Daly, P. Armstrong, M. Szebehely, H. Armstrong, and S. Lafrance, 2012, Structural Violence in Long-Term, Residential Care for Older People: Comparing Canada and Scandinavia, *Social Science & Medicine*, 74(3):390–398.

Barberger-Gateau, P., D. Commenger, M. Gagnon, L. Letenneur, C. Sauvel, and J.-F. Dartigues, 1992, Instrumental Activities of Daily Living as a Screening Tool for Cognitive Impairment and Dementia in Elderly Community Dwellers, *Journal of the American Geriatrics Society*, 40(11):1129–1134.

Becker, U. and H.-J. Reinhard, 2018, *Long-Term Care in Europe: A Juridical Approach.* Springer, Munich.

Bettio, F., A. Simonazzi, and P. Villa, 2006, Change in Care Regimes and Female Migration: the Care Drain in the Mediterranean, *Journal of European Social Policy*, 16(3):271–285.

Bolin, K., B. Lindgren, and P. Lundborg, 2008, Informal and Formal Care Among Single-Living Elderly in Europe, *Health Economics*, 17(3):393–409.

Bonsang, E., 2009, Does Informal Care from Children to their Elderly Parents Substitute for Formal Care in Europe?, *Journal of Health Economics*, 28(1):143–154.

Börsch-Supan, A., 2018a, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1*. SHARE-ERIC.

Börsch-Supan, A., 2018b, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2*. SHARE-ERIC.

Börsch-Supan, A., 2018c, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3*. SHARE-ERIC.

Börsch-Supan, A., 2018d, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4*. SHARE-ERIC.

Börsch-Supan, A., 2018e, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5*. SHARE-ERIC.

Börsch-Supan, A., 2018f, *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6*. SHARE-ERIC.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber, 2013, Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE), *International Journal of Epidemiology*, 42(4):992–1001.

Börsch-Supan, A., K. Hank, and H. Jürges, 2005, A New Comprehensive and International View on Ageing: Introducing the Survey of Health, Ageing and Retirement in Europe, *European Journal of Ageing*, 2(4):245–253.

Brody, E. M. and M. P. Lawton, 2018, *Les Activités Instrumentales de la Vie Quotidienne*. Département de l'Information Médicale.

Carrera, F., E. Pavolini, C. Ranci, and A. Sabbatini, 2013, *Reforms in Long-Term Care Policies in Europe*. Springer New York, New York.

Choi, N. G., 1994, Patterns and Determinants of Social Service Utilization: Comparison of the Childless Elderly and Elderly Parents Living With or Apart From Their Children, *The Gerontologist*, 34(3):353–362.

Comas-Herrera, Adelina, Wittenberg, R., J. Costa-Font, C. Gori, A. Di Maio, C. Patxot, L. Pickard, A. Pozzi, and H. Rothgang, 2006, Future Long-Term Care Expenditure in Germany, Spain, Italy and the United Kingdom, *Ageing and Society*, 26(2):285–302.

Costa-Font, J., 2010, Family Ties and the Crowding Out of Long-Term Care Insurance, *Oxford Review of Economic Policy*, 26(4):691–712.

Costa-Font, J., 2012, *Europe and the Mediterranean Economy*. Routledge, New York.

Costa-Font, J., C. Courbage, and J. Wagner, 2019, Long-Term Care Insurance Research and Trajectory, *The Geneva Papers on Risk and Insurance - Issues and Practice*, 44(2):179–182.

Courbage, C. and J. Costa-Font, 2015, Crowding Out of Long-Term Care Insurance: Evidence From European Expectations Data, *Health Economics*, 24(1):77–88.

Courbage, C., G. Montoliu-Montes, and J. Wagner, 2018, Informal Care, Long-Term Care Insurance and Intra-Family Moral Hazard: Empirical Evidence from Italy and Spain, *Working Paper, University of Lausanne.*

Courbage, C. and M. Plisson, 2012, Financing Long-Term Care in France, In J. Costa-Font and C. Courbage, editors, *Financing Long-Term Care in Europe*, chapter 8, pages 125–150. Palgrave Macmillan, New York.

Courbage, C. and N. Roudaut, 2008, Empirical Evidence on Long-Term Care Insurance Purchase in France, *Geneva Papers on Risk and Insurance: Issues and Practice*, 33(4):645–658.

Courbage, C. and N. Roudaut, 2011, Long-Term Care Insurance: The French Example, *European Geriatric Medicine*, 2(1):22–25.

Courbage, C. and P. Zweifel, 2011, Two-Sided Intergenerational Moral Hazard, Long-Term Care Insurance, and Nursing Home Use, *Journal of Risk and Uncertainty*, 43(1):65–80.

Da Roit, B., B. Le Bihan, and A. Österle, 2007, Long-term Care Policies in Italy, Austria and France: Variations in Cash-for-Care Schemes, *Social Policy & Administration*, 41(6):653–671.

Daniilidou, N. V., C. Economou, D. Zavras, J. Kyriopoulos, and E. Georgoussi, 2003, Health and Social Care in Aging Population: an Integrated Care Institution for the Elderly in Greece, *International Journal of Integrated Care*, 3:1–10.

De Meijer, C., M. Koopmanschap, T. B. D' Uva, and E. van Doorslaer, 2011, Determinants of Long-Term Care Spending: Age, Time to Death or Disability?, *Journal of Health Economics*, 30(2):425–438.

Di Gessa, G. and E. Grundy, 2014, The Relationship Between Active Ageing and Health Using Longitudinal Data from Denmark, France, Italy and England, *Journal of Epidemiology and Community Health*, 68(3):261–267.

Dong, X. S., X. Wang, C. Daw, and K. Ringen, 2011, Chronic Diseases and Functional Limitations Among Older Construction Workers in the United States: A 10-Year Follow-up Study, *Journal of Occupational and Environmental Medicine*, 53(4):372–380.

Eling, M. and O. Ghavibazoo, 2018, Research on Long-Term Care Insurance: Status Quo and Directions for Future Research, *The Geneva Papers on Risk and Insurance - Issues and Practice*, pages 1–54.

Eurostat, 2017, *Key Figures on Europe.* Luxembourg.

Fong, J. H. and J. Feng, 2018, Comparing the Loss of Functional Independence of Older Adults in the U.S. and China, *Archives of Gerontology and Geriatrics*, 74:123–127.

Fong, J. H., A. W. Shao, and M. Sherris, 2015, Multistate Actuarial Models of Functional Disability, *North American Actuarial Journal*, 19(1):41–59.

Fonseca, R. and Y. Zheng, 2011, The Effect of Education on Health, *Working Paper, RAND Labor and Population.*

Fox, P., K. Maslow, and X. Zhang, 1999, Long-Term Care Eligibility Criteria for People with Alzheimer's Disease., *Health Care Financing Review*, 20(4):67–85.

Freedman, V. A. and L. G. Martin, 1998, Understanding Trends in Functional Limitations Among Older Americans, *American Journal of Public Health*, 88(10):1457–1462.

Freedman, V. a. and L. G. Martin, 1999, The Role of Education in Explaining and Forecasting Trends in Functional Limitations among Older Americans, *Demography*, 36(4):461–473.

Fried, L. P., C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, and M. A. McBurnie, 2001, Frailty in Older Adults: Evidence for a Phenotype, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3):146–157.

Fuino, M. and J. Wagner, 2018a, Long-Term Care Models and Dependence Probability Tables by Acuity Level: New Empirical Evidence from Switzerland, *Insurance: Mathematics and Economics*, 81:51–70.

Fuino, M. and J. Wagner, 2018b, Old-Age Care Prevalence in Switzerland: Drivers and Future Development, *European Actuarial Journal*, 8(2):321–362.

Fujisawa, R. and F. Colombo, 2009, *The Long-Term Care Workforce: Overview and Strategies to Adapt Supply to a Growing Demand.* 44. OECD Health Policy Studies, OECD Publishing, Paris.

Fukushima, N., J. Adami, and M. Palme, 2010, *The Swedish Long-Term Care System.* European Network of Economic Policy Research Institutes, Stockholm.

Gentili, E., G. Masiero, and F. Mazzonna, 2017, The Role of Culture in Long-Term Care Arrangement Decisions, *Journal of Economic Behavior & Organization*, 143:186–200.

Germain, C. M., E. Vasquez, J. A. Batsis, and D. R. McQuoid, 2016, Sex, Race and Age Differences in Muscle Strength and Limitations in Community Dwelling Older Adults: Data from the Health and Retirement Survey (HRS), *Archives of Gerontology and Geriatrics*, 65:98–103.

Gori, C., 2008, *Le Riforme Regionali per i non Autosufficienti.* Rome.

Guibert, Q. and F. Planchet, 2017, Non-Parametric Inference of Transition Probabilities Based on Aalen-Johansen Integral Estimators for Semi-Competing Risks Data - Application to LTC Insurance, *Working Paper, Institut de Science Financière et d'Assurances.*

Hallas, C., 2011, *Greece*, volume 65. OECD.

Himes, C. L., 2000, Obesity, Disease and Functional Limitation in Later Life, *Demography*, 37(1):73–82.

Hussein, S. and J. Manthorpe, 2005, An International Review of the Long-Term Care Workforce: Policies and Shortages, *Journal of Aging & Social Policy*, 17(4):75–94.

Iwarsson, S., V. Horstmann, and U. Sonn, 2009, Assessment of Dependence in Daily Activities Combined with a Self-Rating of Difficulty, *Journal of Rehabilitation Medicine*, 41(3):150–156.

Janssen, D., W. Jongen, and P. Schröder-Bäck, 2016, Exploring the Impact of Austerity-Driven Policy Reforms on the Quality of the Long-Term Care Provision for Older People in Belgium and the Netherlands, *Journal of Aging Studies*, 38:92–104.

Jiménez-Martín, S., J. M. Labeaga-Azcona, and C. Vilaplana-Prieto, 2016, Interactions between Private Health and Long-term Care Insurance and the Effects of the Crisis: Evidence for Spain, *Health Economics*, 25(11):159–179.

Jiménez-Martín, S. and C. V. Prieto, 2012, The Trade-Off Between Formal and Informal Care in Spain, *European Journal of Health Economics*, 13(4):461–490.

Karlsson, M., T. Iversen, and H. Øien, 2012, Scandinavian Long-Term Care Financing, In J. Costa-Font and C. Courbage, editors, *Financing Long-Term Care in Europe*, chapter 14, pages 254–278.

Katz, S. and B. A. Jackson, 1963, The Index of ALD: A Standardized Measure of Biological and Psychosocial Function, *JAMA*, 185(12):914–919.

Kempen, G. I., A. M. Myers, and L. E. Powell, 1995, Hierarchical Structure in ADL and IADL: Analytical Assumptions and Applications for Clinicians and Researchers, *Journal of Clinical Epidemiology*, 48(11):1299–1305.

Kingston, A., J. Collerton, K. Davies, J. Bond, L. Robinson, and C. Jagger, 2012, Losing the Ability in Activities of Daily Living in the Oldest Old: A Hierarchic Disability Scale from the Newcastle 85+ Study, *PLoS ONE*, 7(2):1–7.

Kuo, H. K., R. N. Jones, W. P. Milberg, S. Tennstedt, L. Talbot, J. N. Morris, and L. A. Lipsitz, 2005, Effect of Blood Pressure and Diabetes Mellitus on Cognitive and Physical Functions in Older Adults: A Longitudinal Analysis of the Advanced Cognitive Training for Independent and Vital Elderly Cohort, *Journal of the American Geriatrics Society*, 53(7):1154–1161.

Lagergren, M., C. Fagerström, B.-M. Sjölund, J. Berglund, L. Fratiglioni, E. Nordell, E. von Strauss, A. Wimo, and S. Elmståhl, 2016, Horizontal and vertical targeting: a population-based comparison of public eldercare services in urban and rural areas of Sweden, *Aging Clinical and Experimental Research*, 28(1):147–158.

Lang, I., J. Guralnik, R. B. Wallace, and D. Melzer, 2007, What Level of Alcohol Consumption Is Hazardous for Older People? Functioning and Mortality in U.S. and English National Cohorts, *Journal of the American Geriatrics Society*, 55(1):49–57.

Lau, K. M., M. Parikh, D. J. Harvey, C.-J. Huang, and S. T. Farias, 2015, Early Cognitively Based Functional Limitations Predict Loss of Independence in Instrumental Activities of Daily Living in Older Adults, *Journal of the International Neuropsychological Society*, 21(09):688–698.

Lesman-Leegte, I., T. Jaarsma, J. C. Coyne, H. L. Hillege, D. J. Van Veldhuisen, and R. Sanderman, 2009, Quality of Life and Depressive Symptoms in the Elderly: A Comparison Between Patients With Heart Failure and Age- and Gender-Matched Community Controls, *Journal of Cardiac Failure*, 15(1):17–23.

Letenneur, L., V. Gilleron, D. Commenges, C. Helmer, J. M. Orgogozo, and J. F. Dartigues, 1999, Are Sex and Educational Level Independent Predictors of Dementia and Alzheimer's

Disease? Incidence Data from the PAQUID Project, *J Neurol Neurosurg Psychiatry*, 66:177–183.

Mathers, C., 1996, Trends in Health Expectancies in Australia 1981-1993, *Journal of the Australian Population Association*, 13(1):1–15.

Meijer, E., G. Zamarro, K. Arie, and M. Fernandes, 2008, *Comparison Between SHARE, ELSA, and HRS*. Mannheim Research Institute for the Economics of Aging, Mannheim.

Michaud, P.-C. and A. van Soest, 2008, Health and Wealth of Elderly Couples: Causality Tests Using Dynamic Panel Data Models, *Journal of Health Economics*, 27(5):1312–1325.

Mobley, L. R., E. Root, L. Anselin, N. Lozano-Gracia, and J. Koschinsky, 2006, Spatial Analysis of Elderly Access to Primary Care Services, *International Journal of Health Geographics*, 5(19):1–17.

Mols, F., J. W. W. Coebergh, and L. V. van de Poll-Franse, 2007, Health-Related Quality of Life and Health Care Utilisation among Older Long-Term Cancer Survivors: A Population-Based Study, *European Journal of Cancer*, 43(15):2211–2221.

Moritz, D. J., S. V. Kasl, and L. F. Berkman, 1995, Cognitive-Functioning and the Incidence of Limitations in Activities of Daily Living in an Elderly Community Sample, *American Journal of Epidemiology*, 141(1):41–49.

Nadash, P., P. Doty, and M. von Schwanenflügel, 2018, The German Long-Term Care Insurance Program: Evolution and Recent Developments, *The Gerontologist*, 58(3):588–597.

National Institute of Health, 2014, *Parkinson's Disease and Environmental Factors*. U.S. Deptartment of Health and Human Services, Durham.

Navaie-Waliser, M., A. Spriggs, and P. H. Feldman, 2002, Informal Caregiving: Differential Experiences by Gender, *Medical Care*, 40(12):1249–1259.

Nielsen, M., M. Blenkner, M. Bloom, T. Downs, and H. Beggs, 1972, Older Persons After Hospitalization: a Controlled Study of Home Aide Service, *American Journal of Public Health*, 62(8):1094–1101.

Nies, H., K. Leichsenring, and S. Mak, 2013, The Emerging Identity of Long-Term Care Systems in Europe, In K. Leichsenring, J. Billing, and H. Nies, editors, *Long-Term Care in Europe: Improving Policy and Practice*, chapter 2, pages 19–41. Palgrave Macmillan, New York.

OECD, 2013, Long-Term Care in Sweden, In *OECD Reviews of Health Care Quality*, chapter 3, pages 119–170. OECD Publishing.

Österle, A., 2010, Long-term Care in Central and South-Eastern Europe: Challenges and Perspectives in Addressing a New Social Risk, *Social Policy & Administration*, 44(4):461–480.

Österle, A., 2012, Long-Term Care Financing in Central Eastern Europe, In J. Costa-Font and C. Courbage, editors, *Financing Long-Term Care in Europe*, chapter 13, pages 236–237. Palgrave Macmillan, New York.

Pinquart, M. and S. Sörensen, 2011, Spouses, Adult Children, and Children-in-Law as Caregivers of Older Adults: A Meta-Analytic Comparison, *Psychology and Aging*, 26(1):1–14.

Pinquet, J., M. Guillén, and M. Ayuso, 2011, Commitment and Lapse Behavior in Long-Term Insurance: A Case Study, *Journal of Risk and Insurance*, 78(4):983–1002.

Pollack, C. E., S. Chideya, C. Cubbin, B. Williams, M. Dekker, and P. Braveman, 2007, Should Health Studies Measure Wealth?. A Systematic Review, *American Journal of Preventive Medicine*, 33(3):250–264.

Price, M. D., J. F. Hurdle, K. R. Wildes, K. J. O'Malley, C. M. Ashton, and K. F. Cook, 2005, Measuring Diagnoses: ICD Code Accuracy, *Health Research and Educational Trust*, 40(5):1620–1639.

Rubenstein, L. Z., C. Schairer, G. D. Wieland, and R. Kane, 1984, Systematic Biases in Functional Status Assessment of Elderly Adults: Effects of Different Data Sources, *Journals of Gerontology*, 39(6):686–691.

Saraceno, C., 2010, Social Inequalities in Facing Old-Age Dependency: a Bi-Generational Perspective, *Journal of European Social Policy*, 20(1):32–44.

Sinclair, A. J., S. P. Conroy, and A. J. Bayer, 2008, Impact of Diabetes on Physical Function in Older People, *Diabetes Care*, 31(2):233–235.

Sloan, F. A. and E. C. Norton, 1997, Adverse Selection, Bequests, Crowding Out, and Private Demand for Insurance: Evidence from the Long-Term Care Insurance Market, *Journal of Risk and Uncertainty*, 15(3):201–219.

Smith, C. D., H. Chebrolu, D. R. Wekstein, F. A. Schmitt, and W. R. Markesbery, 2007, Age and Gender Effects on Human Brain Anatomy: A Voxel-Based Morphometric Study in Healthy Elderly, *Neurobiology of Aging*, 28(7):1075–1087.

Sowa, A., 2010, *The System of Long-Term Care in the Czech Republic*. European Network of Economic Policy Research Institutes, Warsaw.

Steinbeisser, K., E. Grill, R. Holle, A. Peters, and H. Seidl, 2018, Determinants for Utilization and Transitions of Long-Term Care in Adults 65+ in Germany: Results from the Longitudinal KORA-Age Study, *BMC Geriatrics*, 18(172).

Strauss, E., H. Agüero-Torres, I. Kåreholt, B. Winblad, and L. Fratiglioni, 2003, Women are More Disabled in Basic Activities of Daily Living than Men Only in Very Advanced Ages: A Study on Disability, Morbidity, and Mortality from the Kungsholmen Project, *Journal of Clinical Epidemiology*, 56(7):669–677.

Styczynska, I. and A. Sowa, 2011, *Determinants of Obtaining Formal and Informal LTC across European Countries*. Case Network Studies and Analyses, Warsaw.

Swiss Re, 2014, How Will We Care? Finding Sustainable Long-Term Care Solutions for an Ageing World, *Sigma*, No 5/2014.

Tediosi, F. and S. Gabriele, 2010, *The Long-Term Care System for the Elderly in Italy*. European Network of Economic Policy Research Institutes, Rome.

Tennstedt, S. L., S. L. Crawford, J. B. Mckinlay, and J. B. Mckinlay, 1993, Is Family Care on the Decline? A Longitudinal Investigation of the Substitution of Formal Long-Term Care Services for Informal Care, *The Milbank Quarterly*, 71(4):601–624.

Tuokko, H., R. Frerichs, J. Graham, K. Rockwood, B. Kristjansson, J. Fisk, H. Bergman, A. Kozma, and I. McDowell, 2003, Five-Year Follow-up of Cognitive Impairment With No Dementia, *Archives of Neurology*, 60(4):577–582.

United Nations, 1997, International Standard Classification of Education - ISCED 1997, (November).

Van Houtven, C. H., N. B. Coe, and R. T. Konetzka, 2015, Family Structure and Long-Term Care Insurance Purchase, *Health Economics*, 24:58–73.

Willemé, P., 2010, *The Long-Term Care System for the Elderly in Belgium.* European Network of Economic Policy Research Institutes, Brussels.

World Health Organization, 2000, *Obesity: Preventing and Managing the Global Epidemic.* WHO Technical Report Series, Geneva.

World Health Organization, 2017, *Depression and Other Common Mental Disorders: Global Health Estimates.* WHO Document Production Services, Geneva.

Zarit, S. H., 1996, Behavioral Disturbances of Dementia and Caregiver Issues, *International Psychogeriatrics*, 8(3):263–272.

Zhou-Richter, T., M. J. Browne, and H. Gründl, 2010, Don't They Care? Or, are They Just Unaware? Risk Perception and the Demand for Long-Term Care Insurance, *Journal of Risk and Insurance*, 77(4):715–747.

Zweifel, P., S. Felder, and A. Werblow, 2004, Population Ageing and Health Care Expenditure: New Evidence on the Red Herring, *The Geneva Papers on Risk and Insurance – Issues and Practice*, 29(4):652–666.

Zweifel, P. and W. Struwe, 1996, Long-Term Care Insurance and Bequests as Instruments for Shaping Intergenerational Relationships, *Journal of Risk and Uncertainty*, 12(1):65–76.

Appendix

	Wave 1	Wave 2	Wave 4	Wave 5
Intercept	−7.952 (.906) ***	−8.918 (0.745) ***	−9.711 (.522) ***	−9.551 (.442) ***
Age	0.083 (.009) ***	0.077 (.008) ***	0.086 (.006) ***	0.085 (.005) ***
Gender (baseline: Male)				
Female	−1.187 (.852)	−1.455 (.807) .	−0.844 (.604)	−0.449 (0.509)
Gender × Age	0.019 (.011) .	0.023 (.010) *	0.013 (.008)	0.008 (.007)
Body mass index (baseline: Normal weight)				
Underweight	0.430 (.254) .	0.731 (.210) ***	0.434 (.182) *	0.508 (.162) **
Overweight	0.216 (.088) *	0.201 (.086) *	0.100 (.065)	0.073 (.056)
Moderately obese	0.647 (.112) ***	0.599 (.108) ***	0.612 (.081) ***	0.469 (.070) ***
Severely obese	1.050 (.203) ***	1.121 (.181) ***	1.115 (.133) ***	0.939 (.112) ***
Very severely obese	1.793 (.321) ***	1.534 (.339) **	1.405 (.234) ***	1.584 (.187) ***
Daily smoker (baseline: No)				
Yes	0.036 (.088)	0.103 (.083)	0.019 (.062)	0.169 (.052) **
Partner in household (baseline: No)				
Yes	−0.161 (.089) .	−0.192 (.083)	−0.223 (.063) ***	−0.280 (.053) ***
Children (baseline: No)				
Yes	−0.067 (.107)	0.155 (.111)	−0.013 (.085)	0.041 (.075)
Wealth status (baseline: High)				
Mid-high	−0.038 (.103)	0.091 (.099)	0.366 (.073) ***	0.294 (.060) ***
Mid-low	0.335 (.110) **	0.423 (.107) ***	0.721 (.080) ***	0.602 (.066) ***
Low	0.410 (.144) **	0.690 (.136) ***	1.060 (.107) ***	1.113 (.089) ***
Education level (baseline: Primary)				
Secondary	−0.118 (.087)	−0.085 (.083)	0.054 (.064)	−0.153 (.057) **
Tertiary	−0.287 (.134) *	−0.244 (.128) .	−0.187 (.093) *	−0.331 (.079) ***
Parkinson disease (baseline: No)				
Yes	2.059 (.316) ***	2.364 (.277) ***	2.346 (.201) ***	1.956 (.156) ***
Cancer (baseline: No)				
Yes	0.282 (.133) *	0.283 (.149) .	0.196 (.105) .	0.415 (.084) ***
Musculoskeletal system diseases (baseline: No)				
Yes	1.329 (.163) ***	1.104 (.161) ***	1.253 (.113) ***	1.092 (.104) ***
Other physical diseases (baseline: No)				
Yes	0.479 (.076) ***	0.824 (.074) ***	0.715 (.056) ***	0.732 (.048) ***
LTC scheme (baseline: None)				
State responsibility	−1.247 (.557) *	−0.134 (.293)	0.338 (.118) **	0.243 (.118) *
Family care	−1.081 (.560) .	0.113 (.297)	0.317 (.126) *	0.352 (.122) **
Subsidiary	−0.896 (.555)	0.291 (.289)	0.571 (.114) ***	0.599 (.115) ***
N	6 269	7 274	11 975	17 031

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.10: Results for regression model (1) applied on data from SHARE waves 1, 2, 4 and 5.

	Wave 1			Wave 2			Wave 5		
Intercept	−8.441	(1.754)	***	−7.543	(1.304)	***	−6.719	(.764)	***
Age	0.090	(.013)	***	0.088	(.013)	***	0.084	(.008)	***
Gender (baseline: Male)									
Female	0.278	(.421)		0.889	(.494)	.	0.136	(.281)	
Body mass index (baseline: Normal weight)									
Underweight	1.524	(.570)	**	0.474	(.373)		1.201	(.345)	***
Overweight	−0.168	(.181)		0.038	(.171)		−0.160	(.108)	
Moderately obese	−0.038	(.223)		0.132	(.212)		−0.019	(.132)	
Severely obese	−0.259	(.401)		0.365	(.322)		0.052	(.199)	
Very severely obese	0.518	(.510)		0.185	(.646)		0.405	(.300)	
Daily smoker (baseline: No)									
Yes	0.075	(.189)		0.011	(.169)		−0.069	(.103)	
Partner in household (baseline: No)									
Yes	−0.393	(1.968)		−1.387	(1.731)		−2.993	(1.015)	**
Partner in household × Age	−0.011	(.025)		0.007	(.022)		0.029	(.013)	*
Partner in household × Gender	0.339	(.373)		0.122	(.328)		0.623	(.203)	**
Children (baseline: No)									
Yes	0.205	(.382)		0.233	(.463)		−0.299	(.239)	
Children × Gender	−0.712	(.460)		−0.668	(.524)		−0.125	(.297)	
Wealth status (baseline: High)									
Mid-high	0.085	(.209)		−0.013	(.197)		−0.012	(.118)	
Mid-low	0.092	(.221)		0.298	(.211)		0.157	(.126)	
Low	0.497	(.290)	.	−0.048	(.262)		0.100	(.165)	
Education level (baseline: Primary)									
Secondary	−0.505	(.178)	**	0.196	(.165)		0.263	(.110)	*
Tertiary	−0.454	(.274)	.	0.614	(.260)	*	0.130	(.155)	
Parkinson disease (baseline: No)									
Yes	0.772	(.451)	.	1.031	(.372)	**	0.543	(.216)	*
Cancer (baseline: No)									
Yes	−0.061	(.259)		−0.224	(.276)		−0.267	(.156)	.
Musculoskeletal system diseases (baseline: No)									
Yes	0.371	(.269)		0.634	(.251)	*	0.329	(.096)	***
Other physical diseases (baseline: No)									
Yes	0.404	(.159)	*	0.334	(.152)	*	0.415	(.168)	*
LTC scheme (baseline: None)									
State responsibility	1.429	(1.256)		−0.204	(.564)		0.481	(.242)	*
Family care	0.165	(1.265)		−1.255	(.571)	*	−0.757	(.251)	**
Subsidiary	1.804	(1.252)		−0.120	(.556)		0.611	(.235)	**
N	983			1 092			2 542		

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.11: Results for regression model (3) applied on data from SHARE waves 1, 2 and 5.

Chapter 3

Drivers of Old-Age Dependence and Long-Term Care Usage in Switzerland: a Structural Equation Model Approach

Long-term care (LTC) encompasses a set of services provided to impaired and dependent elderly. To assess the level of the dependence several scales are used, including activities of daily living (ADL), instrumental ADL and functional limitations. Once an elderly fails to perform these activities independently, he or she requires special assistance. Help can be provided as informal care by relatives and as formal care by professionals. The aim of this research is to study individual characteristics that can help to determine the demand of LTC and define a relation between formal and informal care. Our study is based on data from the Swiss Health Survey and focuses on respondents aged over 65 years. We develop a statistical model using the structural equation modeling technique, which allows representing the dependence concept as a latent variable. This hidden dependence variable combines indices linked to limitations in ADL, instrumental ADL and functional limitations. Accounting for causality links between covariates enables us to include the indirect effect of pathologies on the receipt of LTC mediated via dependence. Furthermore, in our model we do not assume a causal relationship between formal and informal care. From our results, we observe a significant impact of pathologies as well as of the social environment on the demand for LTC. The relationship between formal and informal care is found to be both of complementary and substitutional nature.

Note: This is joint work with J. Wagner.

3.1 Introduction

The population is aging in many countries in the world and the speed of the process is expected to increase (Lutz et al., 2008). The dramatic growth of the number of elderly triggers a higher demand for long-term care (LTC), which is defined by the International Labour Office as the "support that is needed by older persons with limited ability to care for themselves due to physical or mental conditions, including chronic diseases and multi-morbidity" (Scheil-Adlung, 2015). Such care can be delivered as informal care by persons from the social network, i.e., partners, children, friends or neighbors, of a dependent elderly. Further, formal LTC embodies services and assistance in personal care and domestic tasks by specifically trained and qualified persons. Formal LTC is associated with higher expenditures and raises financial challenges on the health policies. As a result, informal care is an important alternative to the more costly formal care (Weaver and Weaver, 2014). Therefore, investigating the link between formal and informal care is of particular importance (see also Fuino and Wagner, 2019), whether this relationship is substitutional or complementary.

The present study investigates the relation between formal and informal care in a structural equation modeling (SEM) framework. Instead of relying on one dependence measure, the SEM technique allows us to measure dependence by several scales, which ensures robustness of our results. Thereby, we mediate the effect of medical conditions on the receipt of LTC via dependence. Based on data from the Swiss Health Survey (SHS), we find that strokes, heart attacks, bronchitis and diabetes significantly increase the level of dependence. We observe this effect to a less strong extent in the case of osteoporosis, arthritis and cancer. Further, higher dependence is associated with higher ages and higher deviations from the average body mass index (BMI). Being a woman increases the probability to require LTC while neither the education level nor the monthly income significantly changes the probability. Residents from French-speaking cantons report a higher probability of receiving formal care. A higher number of persons living in the same household decreases the consumption of formal care, while increasing the use of informal care (substitutional effect). However, we do not observe any impact of the presence of children outside of the household. These findings can be put in relationship with a European study based on the SHARE dataset (cf. Fuino et al., 2019). Finally, we report on the positive correlation between formal and informal care that indicates complementarity.

The remainder of the paper is organized as follows: the next section discusses characteristics that can influence the demand for LTC by reviewing the existing literature and presenting the empirical background. We also derive our research hypotheses. Section 3.3 introduces the dataset of Swiss Health Survey, discusses the variables available for the empirical model and provides descriptive statistics. In Section 3.4, we introduce a statistical model based on SEM, outline our estimation results and review the main findings. We conclude in Section 3.5.

3.2 Development of research hypotheses

In the following, we lay out the landscape of existing literature that we use to derive four research hypotheses.

Age has been extensively studied in conjunction with LTC dependence. The decline of the general health is associated with the natural process of aging through the frailty syndrome

(Fried et al., 2001) which triggers impairment and dependence. Research by House et al. (1990) and Rahmqvist (2001) underlines the relationship between age and self-rated health that is significantly associated with dependence. Further, age is one of the key demographic covariates for assessing the utilization of care facilities (see, e.g., Langa et al., 2001a; Michaud and van Soest, 2008; Kim and Lim, 2015; Coe and Van Houtven, 2009; etc.). Another health-related variable is the BMI. While Pi-Sunyer (1991) finds obesity as a major health risk, Wu et al. (2013) also observe that underweight can affect self-rated health. Finally, the impact of diseases on the demand for LTC is widely accepted. We start by mentioning, e.g., Wong et al. (2010) who study how diagnoses trigger the utilization of different forms of formal care. They find significant influence among diagnoses of cancer, mental diseases (Schizophrenia, dementia and Alzheimer's disease) and others. Similar results are derived in the papers by Michaud and van Soest (2008), De Meijer et al. (2011), Langa et al. (2001a), Hanaoka and Norton (2008) and Shapiro and Tate (1985). However, most of these studies consider the direct effect of diseases. We will emphasize on the importance to measure the mediated effect through the dependence status. We state our first conjecture as follows:

Conjecture 1. The indirect effect of physical diseases, the age and the BMI on the utilization of formal and informal care mediated via dependence is significant and positive.

In all of the care literature, the gender of the elderly is found to play a vital role in predicting and explaining the demand for LTC. Evidences suggests to use the gender as an explanatory variable: women tend to live longer and become disabled at higher ages (Barczyk and Kredler, 2018), the type of care provided to women significantly differs from the one provided to men (e.g., Denton, 1997) and informal caregivers are mostly female individuals entailing higher receipt of informal care by their male spouses (Kemper, 1992 and Wagner and Brandt, 2018). Two other essential components are the education and the wealth of the elderly. Rodrigues et al. (2018) and Freedman and Martin (1999) conclude that the use of LTC significantly differs across the accumulated wealth, and Kemper (1992) shows the significant and positive association between monthly income and the probability to receive formal care. Further, the education level is positively associated with the probability of hospitalization in the paper by Weaver and Weaver (2014). Further references that use education and income include papers by De Meijer et al. (2011), Hanaoka and Norton (2008), Huisman et al. (2003), Shapiro and Tate (1985), Tomiak et al. (2000).

Cultural aspects are important drivers for the usage of formal and informal LTC. A number of studies report on the so-called North-South gradient in Europe (Bolin et al., 2008; Bonsang, 2009; Torbica et al., 2015; and Barczyk and Kredler, 2018). For example, Bolin et al. (2008) provide evidence that in Southern European countries informal care plays a very important role, when compared to Central Europe. This is also derived by Fuino et al. (2019). Switzerland has three cultural clusters linked to the spoken language, namely German-, French- and Italian-speaking parts. The presence of this categorization is well covered, e.g., in Fuino and Wagner (2018). Let us stress the paper by Gentili et al. (2017) who observe that French-speaking population is more likely to enter a nursing home with higher dependency level and age, when compared to elderly from the German-speaking part.

Conjecture 2. Demographic variables such as gender, education, income and Swiss language regions have a significant direct effect on the utilization of formal and informal care.

Aspects from the social environment are often used to predict the demand for various types of LTC. In case of informal care, the assistance is mostly provided by spouses (partners) and children, followed by friends and neighbors (Christianson, 1988). The number of such caregivers, as well as their specific characteristics (age, gender, proximity, etc.) are associated with the use of informal LTC. For instance, Bakx et al. (2015) indicate the spouse's characteristics as being significant drivers of professional services in the Netherlands. Balia and Brau (2014) account for the distance to the nearest child and find it relevant to the average number of hours of informal care. In the below Conjecture 4 we also list studies which use caregivers' characteristics in a framework of the instrumental variable approach.

Further, Finlayson (2002) provides a comprehensive and exhaustive list of references for which the relation between the socio-related factors and formal care is investigated. Most of the reviewed studies focus on the number of informal caregivers (typically represented as the number of persons in the household) and various dummy variables, such as living alone, the presence of children, the presence of the spouse in the household and the marital status. For example, Newman et al. (1990) emphasize the relationship of living alone and institutional care. As in the case of informal care, factors like the gender and the age of potential caregivers are also considered. For example, Jette et al. (1995) outline a strong increase in the risk of entering the nursing home, in cases of elderly with a male caregiver, when compared to those with a female available spouse. Fuino et al. (2019) find that the presence of the partner in the household significantly impacts the usage of professional LTC services. The number of persons in the household incorporates the information whether or not the individual has a potential caregiver, regardless whether this is the spouse or a child. Since at higher ages, children mostly live separately from their elderly parents, we account for the presence of children outside of the household.

Conjecture 3. The utilization of formal and informal care strongly depends on the social environment and the household structure. Formal and informal care are negatively, respectively positively, associated with the household size and the presence of children outside household.

The interaction between formal and informal care has been studied extensively during the past decades. The primary question of this domain of research is to find whether both types of care are substitutes or complements. Previous empirical studies outline mixed results. Chappell (1985) compares elderly on their utilization of home-based care using structured interview. The author finds that the relation is complementary and that formal care does not substitute to informal. Further, Langa et al. (2001b) observe that at-home care services increase the number of hours in case elderly need greater social support indicating a complementary relationship. Christianson (1988) reports that the increase of formal services has no significant impact on informal care utilization while using an experimental design methodology. Finally, a small substitutional effect has been shown by Pezzin et al. (1996).

Another stream of research emphasizes on the importance of accounting for endogeneity of informally provided care which may otherwise lead to inconsistent findings. To cope with this issue, several studies used an instrumental variable approach, which was pioneered by Van Houtven and Norton (2004) and Lo Sasso and Johnson (2002). Various information on family structure, particularly children's characteristics are used as instruments: Charles and Sevak (2005) use the proximity, the gender and the marital status of children, Bolin et al. (2008) employ the age of the oldest child, and Bonsang (2009) exploits the proportion of daughters and the distance to the nearest child. In these studies, informal care is found to be to some extent a substitute of formal care. More recent studies (Weaver and Weaver, 2014; Kim and Lim, 2015; Torbica et al., 2015; and Mommaerts, 2018) confirm these results.

Most of the previous studies implicitly assume the causality of one form of care on the other, e.g., the effect of receiving informal care on the utilization of formal care. However, the decision on which type of services to use is made simultaneously and the causal link is not fully obvious. Given the complex nature, and in order to address a potential bias, we let formal and informal care correlate rather than explain each other. This is possible in a framework of structural equation modeling.

Conjecture 4. The usage of formal and informal care is positively correlated.

3.3 Swiss Health Survey data and descriptive statistics

3.3.1 SHS dataset

We base our analysis on the SHS data provided by the Swiss Federal Statistical Office (FSO). The survey has been conducted every five years starting in 1992 yielding six survey datasets. The target population of this survey are individuals aged over 15 permanently living in Switzerland. Imprisoned individuals and those who live in collective households (boarding schools, hospitals, medical-social establishments, convents, hotels and other institutions) are not eligible for an interview. The interview is split into two parts, an interview over-the-phone complemented with a paper- or electronic-based questionnaire. Three techniques are used to ensure the quality of telephone interviews: computer-assisted telephone interviewing (CATI, 96%), computer-assisted personal interviewing (CAPI, $< 1\%$) and telephone interview with a close person (4%).

In the framework of the SHS, the health concept and the quality of life depends on four key aspects: resources and services in the health sector, the lifestyle and health behavior, the natural and man-made environment and the social environment (Swiss Federal Statistical Office, 2019). Therefore, the interviews cover themes such as sociodemographic characteristics (gender, age, marital status, education, income, household structure, etc.), lifestyle and behavioral risks (tobacco and alcohol consumption, gambling, physical activity, etc.), health (body measurements, chronic diseases, physical disorders, etc.) and health services (medical consultations, hospitalizations, etc.).

In our research we use the most recent survey conducted in 2017. It includes responses of 22 134 individuals from the phone-based interview and 18 832 from the written interview. The survey data are complemented by a file containing indices. Indices are generated by FSO and summarize raw questions in a more convenient, consistent and condensed format. We merge the surveys and indices together along each unique identifier. In our study, we focus only on respondents aged 65 years or older, since LTC utilization occurs at higher ages (Balia and Brau, 2014). Restricting the age to be above 65 and below 99 years yields a dataset containing 5 114 observations. Removing all entries with missing values in the below variables leaves us with 4 048 observations.

In the SHS, the level of the dependence is assessed by the number of limitations in ADL, IADL and functional limitations. Five items used as activities of daily living in SHS are similar to the ones used in the original Katz scale (Katz et al., 1963): eating, getting in/out of bed, dressing, toileting and taking a bath or a shower. The SHS scale, however, does not include continence, which is used in the Katz scale. Survey participants are asked how they can perform each of these activities within the following answers: "Yes, without difficulty", "Yes, with some difficulty", "Yes, with great difficulty" and "Not at all".[1] In order to convert the categorical answers to numeric ones, we assign to the answers values of 0, 1/3, 2/3 and 1, respectively. Then, we sum up these values over all the activities and divide by the number of items of the scale, i.e. 5. The resulting ADL variable takes values between 0 and 1 and not only indicates whether or not a respondent is dependent but also, to some extent, the degree of dependency. In a similar manner we construct two variables linked to the limitations in IADL and to the functional limitations. The instrumental activities include eight items: meal preparation, calling, shopping, doing laundry, doing light housework, occasionally doing heavy housework, doing accounting and using public transport. The answers are again coded as above and the variable $IADL$ is constructed by summing up and dividing by eight. Finally, functional limitations encompass limitations in reading a journal or a book, following a conversation with at least two persons, walking without assistance and speaking without difficulty. We construct the FUN variable using these four activities.

We now discuss selected demographical covariates that we consider in our study. In the SHS the age AGE is represented as an integer number of years in 2017, while the gender SEX can take two values, male and female. Further, the survey contains information on individuals' monthly income INC in CHF. The value does not include compulsory contributions to social insurance but takes into account supportive payments made or received. The question on education allows for 20 different answers for the highest completed level. These answers are summarized in EDU, a compact variable along three categories including primary, secondary and tertiary education levels. Finally, the language of the interview is a good proxy of the linguistic region LNG of the respondent, namely, German-, French- and Italian-speaking regions. Finally, the survey contains information on the height and the weight making it possible to compute the body mass index (BMI). The distribution of the BMI has a bell shape and absolute deviations from the average BMI are associated with a poorer health status. Therefore, we consider absolute deviations $dBMI$ from the sample mean. To reflect the structure of the household we use the number of persons in the household NPR counting the actual persons living in the same household more than four days per week. Furthermore, as respondents are asked whether or not they have children outside the household, we include that number in CHO.

The set of pathologies contained in the SHS includes the following diseases: asthma AST, arthritis ART, osteoporosis OST, bronchitis BRH, heart attack ATT, stroke STR, cancer CNR and diabetes DBT. The binary variables representing these diseases have two levels, yes

[1] The prevalence rate, which is defined as the number of dependent elderly over the total number of elderly, is lower when compared to the one estimated using the SHARE dataset (7.89% versus 9.11%). The lower ratio can be explained by several factors: (1) the ADL definition in SHS data does not include walking activity, (2) the formulation of questions are different in SHARE and SHS, having more detailed answers in SHS, and (3) the underlying population is not the same for these surveys.

and no, that define whether a respondent has the illness or problems related to it during the past 12 months. Finally, the utilization of formal FCR and informal ICR care are represented as binary variables indicating whether the respondent has used services over the past 12 months. Formal care is defined as care services provided at home, such as nursing care, help with domestic tasks, personal care, meals-on-wheels and transport services. Informal care covers assistance in physical care, shopping, preparing meals and administrative tasks provided by a spouse or partner, relatives, neighbors or friends.

3.3.2 Descriptive statistics

In this section, descriptive statistics of the respondents' characteristics are presented. Our first conjecture assumes that the effect of diseases, age and BMI on the utilization of LTC is mediated via dependence. Furthermore, the second and the third conjecture focus on the direct effect of individuals' demographic characteristics as well as the social environment on the receipt of LTC. Therefore, we split exogenous variables into two subgroups depending on the factor they are affecting directly, i.e. dependence or the utilization of LTC.

Exogenous characteristics affecting dependence We start with the variables that directly affect dependence. The descriptive statistics are given in Table 3.1. In each group we provide the number of respondents N and their share within the total data, for each scale that represents the dependence, we report the conditional average value of the dependence measure considering only non-zero values. Overall values of sample conditional means (see bottom of the table) are 0.21, 0.17 and 0.14 for ADL, $IADL$ and FUN, respectively. Among the 4 026 respondents, 292 individuals reported a non-zero value of ADL, while the numbers of individuals reporting non-zero values of $IADL$ and FUN are 1 165 and 970, respectively. Both last numbers are higher, since ADL represent activities that are necessary for fundamental functioning and therefore dependence in ADL implies the dependence in less severe IADL and functional limitations (e.g., Spector et al., 1987 finds that the items "shopping" and "transportation" in IADL show lower levels of dependence when compared to ADL). The three indicators are well related to dependence and highly correlate among each others. Table 3.2 reports information on the correlation and standard deviations of ADL, $IADL$ and FUN.

The overall dataset mainly consists of respondents aged below 85 (93%). The ADL variable does not show a perfect increasing trend for age groups, which can be explained by the fact that which a higher number of limitations, elderly enter nursing homes and are no more represented in the SHS. Further, $IADL$ and FUN increase with age. We group respondents according to their deviations from the sample BMI mean. To form four classes we use sample quantiles, that are at deviations of 1.31 (25%), 2.64 (50%) and 4.64 (75%). We find that dependency increases with BMI deviation. For the group with the largest deviations (4.65+), the averages of the scales are larger when compared to groups with lower deviations.

Finally, we cover the statistics along selected diseases. Arthritis has the highest number of reported cases, which equals to 1 411 and corresponds to 35.0% of the sample. In contrast, the lowest proportion of diseased (1.0%) has had a heart attack (42 cases). This disease comes with the highest value of average ADL (0.43). Such differences can be explained by severe rehabilitation measures after myocardial infarction (Witt et al., 2004). Heart attack is followed by bronchitis, asthma, diabetes and stroke, which also show an important increase in limitations

	N	(%)	ADL	$IADL$	FUN
Age					
65 – 69	1 282	(31.8)	0.20	0.16	0.13
70 – 74	1 145	(28.4)	0.18	0.13	0.13
75 – 79	832	(20.7)	0.24	0.14	0.14
80 – 84	483	(12.0)	0.15	0.16	0.14
85 – 89	218	(5.4)	0.26	0.25	0.20
90 – 94	53	(1.3)	0.22	0.29	0.19
95 – 99	13	(0.3)	0.27	0.43	0.19
BMI deviation					
0.00 – 1.30	1 004	(24.9)	0.19	0.15	0.15
1.31 – 2.62	1 009	(25.1)	0.25	0.16	0.13
2.63 – 4.64	1 005	(25.0)	0.17	0.14	0.13
4.65+	1 008	(25.0)	0.22	0.20	0.16
Asthma					
No	3 841	(95.4)	0.20	0.17	0.14
Yes	185	(4.6)	0.23	0.16	0.15
Arthritis					
No	2 615	(65.0)	0.23	0.17	0.13
Yes	1 411	(35.0)	0.19	0.16	0.15
Osteoporosis					
No	3 639	(90.4)	0.21	0.16	0.14
Yes	387	(9.6)	0.20	0.19	0.16
Bronchitis					
No	3 833	(95.2)	0.20	0.16	0.14
Yes	193	(4.8)	0.24	0.23	0.17
Heart attack					
No	3 984	(99.0)	0.20	0.17	0.14
Yes	42	(1.0)	0.43	0.22	0.19
Stroke					
No	3 983	(98.9)	0.21	0.16	0.14
Yes	43	(1.1)	0.22	0.35	0.22
Cancer					
No	3 851	(95.7)	0.21	0.17	0.14
Yes	175	(4.3)	0.15	0.16	0.12
Diabetes					
No	3 652	(90.7)	0.20	0.16	0.14
Yes	374	(9.3)	0.23	0.23	0.18
Total			0.21	0.17	0.14
N	4 026	(100.0)	292	1 165	970

Table 3.1: Descriptive statistics of exogenous characteristics affecting dependence.

among respondents who reported such diseases compared to those who did not. Cancer, arthritis and osteoporosis have a lower ADL index. The only diseases that lead to a decreasing $IADL$ are asthma, arthritis and cancer. However, the highest increase is shown by stroke. This can be explained by the fact that stroke leads to a strong cognitive decline or even dementia associated with limitations in instrumental activities, especially in those that require cognitive abilities (Brainin et al., 2015). Similar results are obtained for FUN. Both cardio-vascular diseases,

	ADL	$IADL$	FUN
ADL	1.00	0.67	0.49
$IADL$	0.67	1.00	0.63
FUN	0.49	0.63	1.00
Std. dev.	0.08	0.14	0.08

Table 3.2: Correlation coefficients and standard deviations of the indicator variables for dependence.

heart attack and stroke, yield the highest value of FUN. Overall, we conclude that most diseases substantially increase the dependence indicators.

Exogenous characteristics affecting LTC usage The second set of exogenous covariates directly determines LTC usage. In addition to gender, education, income, language group and the household composition, we consider three variables that describe the state of dependence, i.e. ADL, $IADL$ and FUN. For each of these exogenous variables we provide the shares of respondents who reported formal and informal care utilization. These figures are shown in Table 3.3. Among the 4 026 observations, 6.6% of the respondents have used formal care, while more than double of this share (13.7%) have required informal assistance.

Our dataset is composed of 48.3% of males and 51.7% of females. Females show almost twice as high percentage values of use of formal (8.5%) and informal care (17.2%) when compared to the male respondents (4.4% and 10.0%, respectively). About half of the sample has completed secondary education, while the rest of respondents are equally split between primary and tertiary levels. We note that for both formal and informal care the primary education has the highest share of reported use of LTC, followed by secondary and tertiary. In our sample, the income spans from zero to CHF 150 000. To categorize this variable, we use sample quantiles at 25%, 50% and 75%, that are CHF 1 750, 3 000 and 5 000, respectively. We remark that the use of informal care decreases in the monthly income. This pattern does not hold for formal care. The majority of individuals have responded to the questions in German (2 778), followed by French (925) and Italian (323). For informal care, the Italian-speaking respondents indicate a higher use of informal care (18.9% versus 12.8% and 13.5%).

The number of persons in the household ranges from one to nine and most respondents live in two persons household (63.6%). The number of respondents living in households with three or more cohabitants is 276. We see the most important differences in the distributions when considering the number of persons in the household, especially for formal care. In the case where people live alone, 12.5% of the sample use professional assistance. This number decreases to 4.1% and 4.0% in case of two and three and more persons living in the same household. A similar drop can be found in the case of informal care in two persons household. Further, we conclude that most of the respondents have children outside the household (82.4%). The presence of children outside the household is relevant only for professional assistance, when children help to avoid formal care usage.

Finally, we note that the utilization of both care types importantly increases with the ADL, $IADL$ and FUN scale. Throughout the statistics, this effect is expected and consistent. The

	N	(%)	Formal care without (%)	Formal care with (%)	Informal care without (%)	Informal care with (%)
Gender						
Male	1 943	(48.3)	95.6	4.4	90.0	10.0
Female	2 083	(51.7)	91.5	8.5	82.8	17.2
Education						
Primary	905	(22.5)	91.0	9.0	83.1	16.9
Secondary	2 126	(52.8)	93.7	6.3	86.8	13.2
Tertiary	995	(24.7)	95.1	4.9	87.9	12.1
Income (in CHF)						
0 – 1 749	995	(24.7)	94.0	6.0	83.5	16.5
1 750 – 2 999	1 013	(25.2)	91.8	8.2	85.1	14.9
3 000 – 4 999	1 008	(25.0)	94.2	5.8	87.5	12.5
5 000+	1 010	(25.1)	93.8	6.2	88.9	11.1
Language region						
German	2 778	(69.0)	94.3	5.7	86.5	13.5
French	925	(23.0)	91.1	8.9	87.2	12.8
Italian	323	(8.0)	92.3	7.7	81.1	18.9
Number of persons in household						
1	1 189	(29.5)	87.5	12.5	84.4	15.6
2	2 561	(63.6)	95.9	4.1	87.4	12.6
3+	276	(6.9)	96.0	4.0	83.3	16.7
Children (outside the household)						
Yes	3 316	(82.4)	93.8	6.2	86.2	13.8
No	710	(17.6)	91.8	8.2	86.8	13.2
ADL scale (ADL)						
0	3 734	(92.7)	95.8	4.2	89.7	10.3
0.000 – 0.249	215	(5.3)	69.3	30.7	49.8	50.2
0.250 – 0.499	49	(1.2)	49.0	51.0	18.4	81.6
0.500 – 0.749	14	(0.3)	35.7	64.3	0.0	100.0
0.750 – 1.000	14	(0.3)	57.1	42.9	50.0	50.0
IADL scale ($IADL$)						
0	2 861	(71.1)	97.9	2.1	94.8	5.2
0.001 – 0.249	935	(23.2)	88.6	11.4	75.6	24.4
0.250 – 0.499	121	(3.0)	65.3	34.7	34.7	65.3
0.500 – 0.749	51	(1.3)	43.1	56.9	17.6	82.4
0.750 – 1.000	58	(1.4)	56.9	43.1	6.9	93.1
Functional limitations scale (FUN)						
0	3 056	(75.9)	96.1	3.9	90.5	9.5
0.001 – 0.249	803	(19.9)	89.3	10.7	80.3	19.7
0.250 – 0.499	147	(3.7)	64.6	35.4	40.8	59.2
0.500 – 0.749	15	(0.4)	60.0	40.0	0.0	100.0
0.750 – 1.000	5	(0.1)	60.0	40.0	20.0	80.0
Total	4 026	100.0	93.4	6.6	86.3	13.7

Table 3.3: Descriptive statistics of exogenous characteristics affecting LTC usage.

smaller number of observations (less than 15) in classes with higher limitations may explain slight deviations. We observe that increase is less rapid in $IADL$ when compared to the effect of ADL. The functional limitations impact the consumption of formal and informal care less than the other two measures.

Both types of care are found to be used side-by-side in a complementary way by elderly. Table 3.4 reports the confusion matrix of informal and formal care.

	Formal care		
Informal care	No	Yes	Total
No	3 354	119	3 473
Yes	408	145	553
Total	3 762	264	4 026

Table 3.4: Confusion matrix of informal and formal care.

3.4 Model setup, results and discussion

3.4.1 Structural Equation Model

In our study, we base our model on the framework of SEM. The choice of this model is driven by several methodological requirements. First, the concept of dependence cannot be measured directly by a unique single measure. However, it is well represented as a latent (hidden) variable using a factor analysis approach linked to dependency indicators. Further, our first conjecture assumes the effect of physical characteristics on the receipt of LTC. The link, however, between the usage of LTC and physical characteristics is not direct but pipelined through the status of dependence. In addition, these links are assumed to have directions of causality that can be modeled in a framework of path analysis. The combination of factor and path analysis embodies the principles of SEM (see, e.g., Beran and Violato, 2010 and Nachtigall et al., 2003). Despite the high level of utilization of SEM in the clinical, behavioral psychology, health sciences and epidemiology literature, it is less used in economics contexts. Nevertheless, several studies by Baranoff et al. (2007), Lu et al. (2012), Eling and Marek (2013) and references therein have applied SEM in the insurance context.[2]

Figure 3.1 presents the SEM used in the sequel. In the following we discuss the three major parts of the model, namely, the measurement of dependence, the regression of physical characteristics for dependence and, finally, the regression for which endogenous variables are the usage of formal and informal care. Table 3.5 summarizes the variables included in the model.

Measurement of dependence We describe how the concept of dependence is modeled as a latent unobservable variable. The latent variable is measured by indicators (so-called manifest variables), which are assumed to perfectly correlate with the underlying concept. We select three indicators to assess the dependence: the measure of limitations in ADL (ADL), the limitations

[2]In addition, in our context we would not be able to use any survival models, since the SHS data are not longitudinal.

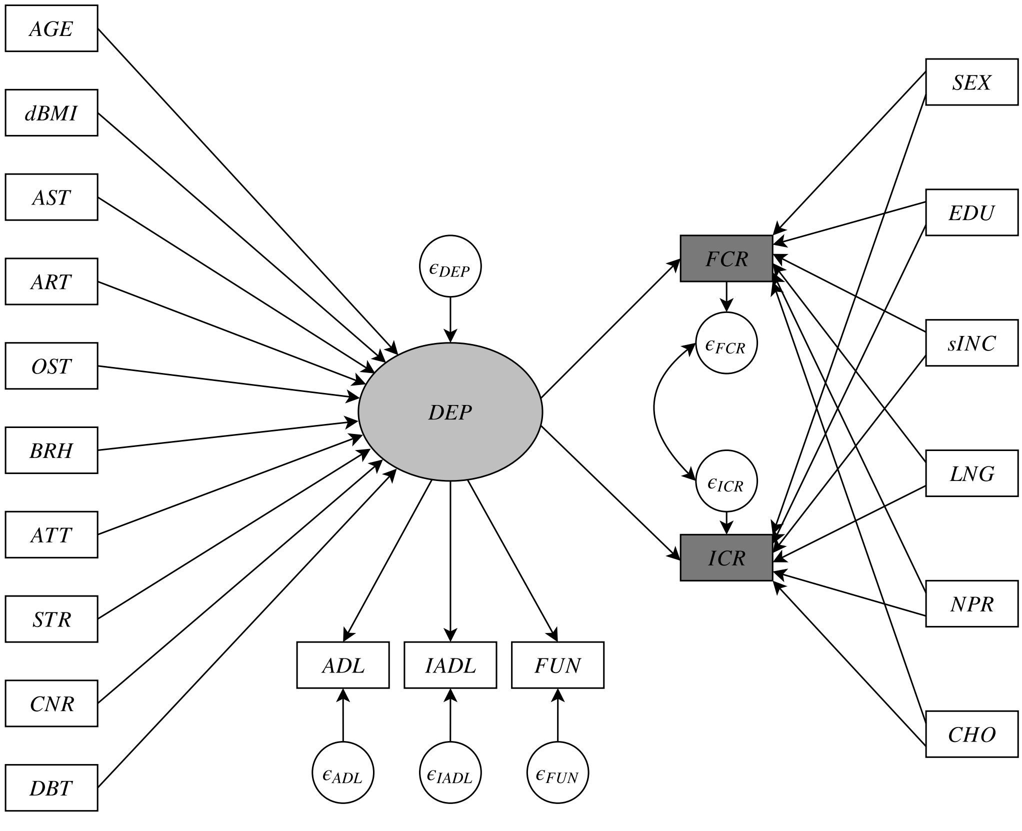

Figure 3.1: Graphical representation of the proposed SEM.

in IADL ($IADL$) and functional limitations (FUN). The system of equations that represents the measurement model of dependence, can be formulated as follows:

$$\begin{aligned} ADL_i &= \lambda_{ADL} DEP_i + \epsilon_{ADL,i}, \\ IADL_i &= \lambda_{IADL} DEP_i + \epsilon_{IADL,i}, \\ FUN_i &= \lambda_{FUN} DEP_i + \epsilon_{FUN,i}, \end{aligned} \tag{5}$$

where λ_{ADL}, λ_{IADL} and λ_{FUN} are factor loadings corresponding to the magnitude of expected change in the observed variable for a one unit change in the latent variable. The terms $\epsilon_{ADL,i}$, $\epsilon_{IADL,i}$ and $\epsilon_{FUN,i}$ are error terms for the individual i. The error terms are assumed to be uncorrelated with each other, with the latent variable DEP, and to have an expected value of zero.

Regression model for dependence Further, we describe the simultaneous equations, which determine the interdependence of the variables in our model. First, we determine the relationship between the age, the BMI and the pathologies with the latent variable of dependence:

$$\begin{aligned} DEP_i = \beta_0 + \beta_{AGE} AGE_i + \beta_{dBMI} dBMI_i + \beta_{AST} AST_i + \beta_{ART} ART_i + \beta_{OST} OST_i + \\ \beta_{BRH} BRH_i + \beta_{ATT} ATT_i + \beta_{STR} STR_i + \beta_{CNR} CNR_i + \beta_{DBT} DBT_i + \epsilon_{DEP,i}. \end{aligned} \tag{6}$$

Here AGE_i is the respondent's age, $dBMI_i$ is the absolute value of deviation from the average BMI, and AST_i, ART_i, OST_i, BRH_i, ATT_i, STR_i, CNR_i, DBT_i are binary variables taking the value of one if the respondent has been diagnosed with asthma, arthritis, osteoporosis, bronchitis, heart attack, stroke, cancer and diabetes, respectively, and zero otherwise. We denote the regression coefficients by β with respective subscripts according to the independent variables. The intercept is denoted by β_0. The error term expressed through $\epsilon_{DEP,i}$ is assumed to have a zero expected value, a constant variance and not to be correlated with the other error terms.

Variables	Type	Description	Values
AGE	Exogenous	Age	from 65 to 99
$dBMI$	Exogenous	Deviations from BMI	from 0 to 42.4
AST	Exogenous	Asthma	yes, no
ART	Exogenous	Arthritis	yes, no
OST	Exogenous	Osteoporosis	yes, no
BRH	Exogenous	Bronchitis	yes, no
ATT	Exogenous	Heart attack	yes, no
STR	Exogenous	Stroke	yes, no
CNR	Exogenous	Cancer	yes, no
DBT	Exogenous	Diabetes	yes, no
DEP	Latent	Dependence level	–
ADL	Manifest	Level of ADL limitations	from 0 to 1
$IADL$	Manifest	Level of IADL limitations	from 0 to 1
FUN	Manifest	Level of functional limitations	from 0 to 1
SEX	Exogenous	Gender	male, female
EDU	Exogenous	Education level	primary, secondary, tertiary
$sINC$	Exogenous	Standardized monthly income	from 0.8 to 33.1
LNG	Exogenous	Language region	German, French, Italian
NPR	Exogenous	Number of persons in the household	from 0 to 9
CHO	Exogenous	Children outside of the household	yes, no
FCR	Endogenous	Use of formal care	yes, no
ICR	Endogenous	Use of informal care	yes, no

Table 3.5: Description of the variables used in the model.

Regression model for LTC usage Further two equations express the dependence of receipt of formal and informal care on socio-demographic characteristics and dependence. We define FCR and ICR as binary variables that take a value of one, if the respondent reported receiving formal respectively informal care, and zero otherwise. We use an unobserved-variable formulation of *probit* regression model, which assumes that there exist underlying continuous unobserved variables FCR^* and ICR^* that are defined as follows:

$$\begin{aligned} FCR_i^* = \gamma_0^{FCR} + \gamma_{DEP}^{FCR} \cdot DEP_i + \gamma_{SEX}^{FCR} \cdot SEX_i + \gamma_{EDU}^{FCR} \cdot EDU_i + \gamma_{sINC}^{FCR} \cdot sINC_i + \\ \gamma_{LNG}^{FCR} \cdot LNG_i + \gamma_{NPR}^{FCR} \cdot NPR_i + \gamma_{CHO}^{FCR} \cdot CHO_i + \epsilon_{FCR,i}, \end{aligned} \tag{7}$$

and

$$\begin{aligned} ICR_i^* = \gamma_0^{ICR} + \gamma_{DEP}^{ICR} \cdot DEP_i + \gamma_{SEX}^{ICR} \cdot SEX_i + \gamma_{EDU}^{ICR} \cdot EDU_i + \gamma_{sINC}^{ICR} \cdot sINC_i + \\ \gamma_{LNG}^{ICR} \cdot LNG_i + \gamma_{NPR}^{ICR} \cdot NPR_i + \gamma_{CHO}^{FCR} \cdot CHO_i + \epsilon_{ICR,i}, \end{aligned} \tag{8}$$

where SEX_i, EDU_i, $sINC_i$, LNG_i, NPR_i and CHO_i are the gender, education level, standardized monthly income,[3] language region, number of persons in household and presence of children outside household, respectively. The γ denote the regression coefficients with γ_0^{FCR} and γ_0^{ICR} being the model intercepts. Subscripts of γ determine the corresponding independent variable and superscripts determine the type of care of the coefficient. We impose assumptions that $\epsilon_{FCR,i} \sim \mathcal{N}(0,1)$, $\epsilon_{ICR,i} \sim \mathcal{N}(0,1)$ and allow $\epsilon_{FCR,i}$ and $\epsilon_{ICR,i}$ to correlate. The observed FCR_i and ICR_i are then defined as $FCR_i = 1$, if $FCR_i^* > 0$, and 0 otherwise, and $ICR_i = 1$, if $ICR_i^* > 0$, and 0 otherwise.

3.4.2 Results and discussion

Our SEM results are shown in Tables 3.6 to 3.8, where we report the factor loadings for the measurement of the latent dependence variable, regression coefficients estimates, standard errors and their statistical significance. By convention, all coefficients are standardized, i.e., each coefficient is multiplied by the ratio of the standard deviations of the explanatory and the dependent variables. This leads to the following interpretation: the standardized regression coefficients represent the expected change of the dependent variable in standard deviation units per standard deviation unit change in the explanatory variable while holding all other explanatory variables constant. We also do not report the intercept, since by convention deviations from the mean are considered. We denote statistical significance by "." when a p-value below 0.1, "*" – below 0.05, "**" – below 0.01 and "***" – below 0.001.

Our model is fitted using a three-stage diagonally weighted least squares estimation, which does not require the assumption of the multivariate normal distribution of the endogenous variables in contrast to maximum likelihood (Muthén, 1993). We calibrate our model by utilizing the R package *lavaan* (Rosseel, 2012). Our model has 68 degrees of freedom. The fitted model has the following goodness-of-fit measures: Tucker Lewis Index of 0.998, Comparative Fit Index of 0.984, Root Mean Square Error of Approximation of 0.034 and Standardized Root Mean Square Residual of 0.024. According to the cut-off points (Hooper et al., 2008), these measures indicate a good fit. We intentionally do not present chi-square p-value because it does not tolerate violation of multivariate normality and would reject the model even if it were correctly specified. Further, our sample is considered as rather large which may also lead to false rejections (see Hooper et al., 2008 and the references therein for details on SEM goodness-of-fit).

Measurement of dependence First, we discuss the measurement of dependence. We choose to set the scale of the latent variable DEP to its first indicator ADL in Equation 5, which is the level of limitations in ADL. When the level of limitations in ADL is high, this scale assumes high values of DEP associated with a higher level of dependence. We report factor loadings as well as their standard deviation and statistical significance in Table 3.6. The factor loadings of the other two indicators, the level of limitations with IADL and functional limitations, have positive and significant values of 2.312 and 0.964, implying that three variables reliably measure the latent construct.

[3]The values of INC range from zero to CHF 150 000 which yields a high scale of variance, when compared to other variables. In order to attain better robustness while calibrating the model, we standardize the income $sINC$, that is $sINC = (INC - \overline{INC})/s_{INC}$, where $\overline{INC}$ and s_{INC} are sample mean and sample standard deviations of income, respectively.

	Dependence measurement (5)		
	λ	Std. dev.	Sig.
$ADL_i \sim \lambda_{ADL} DEP_i$	1.000		
$IADL_i \sim \lambda_{IADL} DEP_i$	2.312	(0.027)	***
$FUN_i \sim \lambda_{FUN} DEP_i$	0.964	(0.012)	***

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.6: Results for measurement of dependence (model 5).

Regression model for dependence We now highlight the regression model (6) for which the dependence is the outcome variable. The coefficient estimates, standard deviations and their significance are reported in Table 3.7. Expectedly, the impact of the age on the dependence is significant and positive, meaning that individuals are expected to have higher levels of dependence at higher ages. Indeed, as mentioned before, higher ages trigger the frailty syndrome associated with poorer health implying strong impairment. The deviation from normal BMI is also found to affect the state of dependence significantly. The positive value of the coefficient indicates that for one standard deviation increase of BMI deviation, the value of the hidden variable of dependence will increase by 0.001 standard deviations. Deviations from the normal BMI are typically considered as indication of severe diseases. For instance, a low BMI is a risk factor for fractures (De Laet et al., 2005) while an increased BMI is associated with an increased risk for various types of cancer (Renehan et al., 2008), cardio-vascular diseases (Gregg et al., 2005) and stroke (Kurth et al., 2002 and Rexrode et al., 1997).

All considered diseases, except asthma and cancer, have a statistically significant impact on dependence. Headen Jr. (1993) finds that asthma decreases the hazard of entering the nursing home by 25% when compared to those who do not have a respiratory disease. At the same time, respiratory diseases increase the hazard of death by 25%. This result allows us to assume that chronic asthma skips the state of dependence leading directly to death of an individual. We note that stroke, bronchitis and heart attack show the highest coefficients. Stroke is found to be the major factor of entering a nursing home after mental diseases (Tomiak et al., 2000). The aftermath of cardio-vascular diseases is typically associated with a long rehabilitation process among survivors (Torbica et al., 2015) and, therefore, yield higher utilization of care services. We could not find similar results in the case of bronchitis which is typically combined with other chronic obstructive pulmonary diseases (COPD). Since many researchers mistakenly also consider asthma as COPD and combine it together with bronchitis, it is hard to distinguish the separate effect of these pathologies.

Individuals that are diagnosed with diabetes have a higher risk of becoming dependent. According to work by Charles and Sevak (2005) who also found diabetes to increase the probability of entering the nursing home, diabetes is associated with a severe condition of heart illness. Further, even though arthritis and osteoporosis are two different diseases with different diagnoses and symptoms, in the context of the dependence, they show similar results. Finally, we report that cancer does not affect the dependence status significantly. The explanation is similar to asthma: cancer typically leads to death faster than other diseases coming with a lower utilization of care services.

	Dependence (6)		
	β	Std. dev.	Sig.
Age			
	0.002	(0.0001)	***
BMI deviations			
	0.001	(0.0003)	***
Asthma (baseline: No)			
Yes	0.000	(0.0046)	
Arthritis (baseline: No)			
Yes	0.008	(0.0021)	***
Osteoporosis (baseline: No)			
Yes	0.012	(0.0028)	***
Bronchitis (baseline: No)			
Yes	0.035	(0.0034)	***
Heart attack (baseline: No)			
Yes	0.034	(0.0055)	***
Stroke (baseline: No)			
Yes	0.071	(0.0047)	***
Cancer (baseline: No)			
Yes	0.007	(0.0049)	
Diabetes (baseline: No)			
Yes	0.020	(0.0026)	***

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.7: Results for the regression model for dependence (model 6).

Regression models for LTC usage When analyzing the drivers of demand for formal and informal LTC, we observe a significant association with dependence with this association being stronger for informal care. Further, as previous studies consistently observed, we find a significant effect of the gender variable: female respondents tend to have a higher probability to receive LTC. This can be explained by the lower mortality and higher frailty at high ages among women. Neither education nor income plays significant role in explaining the demand of long-term care. At first glance, this result might look confusing because both education and income are indicators of socio-economic status and therefore can be proxies of the profession. However, the risk related to the profession of the respondent is also likely to be incorporated in the observed chronic diseases that enter our model via dependence. Another possible explanation is that in Switzerland about 65% of costs related to health care are provided by compulsory health insurance and only 20% are paid out-of-pocket (Gentili et al., 2017). This holds true in particular at-home based care. Therefore, there is a little incentive to save money by substituting formal care by informal care.

The respondents who completed the survey in French language show significantly higher receipt of formal care and lower usage of informal care. This result coincides with the conclusion made by Gentili et al. (2017) who highlight differences in German- and French-speaking parts of Switzerland. Their main finding is that the French-speaking population enters nursing homes only with higher dependence status. As our sample does not contain elderly residing in nursing homes, respondents of the German-speaking part probably have less severe dependence and therefore trigger less use of formal and informal care.

	Formal care (7)			Informal care (8)		
	γ	Std. dev.	Sig.	γ	Std. dev.	Sig.
Dependence						
	6.498	(0.277)	***	8.304	(0.238)	***
Gender (baseline: Male)						
Female	0.232	(0.079)	**	0.350	(0.059)	***
Education (baseline: Primary)						
Secondary	−0.028	(0.081)		0.000	(0.066)	
Tertiary	0.019	(0.106)		0.058	(0.083)	
Income						
	−0.014	(0.043)		0.030	(0.023)	
Language region (baseline: German)						
French	0.223	(0.079)	**	−0.110	(0.065)	.
Italian	0.051	(0.127)		0.133	(0.091)	
Number of persons in household						
	−0.298	(0.056)	***	0.108	(0.038)	**
Children (outside household) (baseline: No)						
Yes	−0.072	(0.084)		0.044	(0.069)	

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.8: Results for the regression models for LTC usage (models 7 and 8).

Finally, we briefly discuss the socio-related variables. The presence of children outside the household does not play a relevant role. We find that a higher number of persons in the household reduces formal care usage. The increase of the number of persons in household by one standard deviation triggers 0.298 decrease of standard deviation of formal care. At the same time, we observe a 0.108 increase in informal care. The opposite signs of these coefficients indicate a substitutional relationship between both types of care: the decrease in formal care is complemented with an increase of informal care delivered by household cohabitants. Overall, we observe a positive and significant correlation coefficient (0.441) between the error terms ϵ_{FCR} and ϵ_{ICR} indicating complementary usage.

To summarize this section, we put the observations in relation with our initial conjectures. By showing the significant effects of pathologies, age and BMI on dependence as well as the significance of dependence on both formal and informal care, we provide the necessary evidence to support the first conjecture. Further, we partially validate the second hypothesis: the gender and the linguistic regions are proven to significantly affect the demand for LTC. Education and income are not directly significant in models (7) and (8). Further, while the presence of children outside of the household does not have a significant effect, the number of persons in the household impacts the demand for both types of LTC. We find that cohabitants act as informal caregivers reducing the demand for formal care. This yields to partial validation of conjecture 3. Finally, our findings support the hypothesis of positive correlation between formal and informal LTC.

3.5 Concluding remarks

In this paper, we investigate and analyze the factors associated with the usage of formal and informal LTC. We develop a statistical model in a framework of structural equation modeling and calibrate it using the data of 4 026 observations from the Swiss Health Survey.

We model the concept of dependence as a latent variable with three manifest indicators, namely limitations in ADL, instrumental ADL and functional limitations. We allow for a direct effect of age, BMI and selected diseases on dependence which, in turn, is affecting the receipt of LTC. In other words, the effect of such factors is mediated via dependence. In addition, we account for demographic and socio-related factors as drivers of demand for LTC. Finally, we let formal and informal care correlate to investigate the nature of the relation.

Our main findings can be summarized as follows: (1) the indirect effects of age, BMI and pathologies on the receipt of LTC are significant while mediated via latent variable of dependence, (2) demographic and socio-environmental variables as the gender, the linguistic region and the number of persons in the household plays a significant role in determining the needs LTC and (3) the relation of formal and informal care can be both substitutional and complementary depending on a context.

Our observations also underline the importance to incorporate the path analysis instead of directly assuming a causal link between formal and informal care. We also add to the understanding of the relationship between formal and informal care. This paper reinforces the importance of medical characteristics of elderly, when studying the evolution of needs of impaired elderly. Finally, our results give indications and insights for the rapid development of new LTC insurance products. The purpose of this paper is to explain the factors that affect LTC use, rather than to predict the demand for LTC. Therefore, the study is limited in potential applications to the process of insurance products' pricing (e.g., due to the choice of methodology and the bias of the sample). However, our research can be a basis for an underwriting questionnaire to assess the risk presented by new customers.

Bibliography

Bakx, P., C. De Meijer, F. Schut, and E. van Doorslaer, 2015, Going Formal or Informal, Who Cares? The Influence of Public Long-Term Care Insurance, *Health Economics*, 24(6):631–643.

Balia, S. and R. Brau, 2014, A Country for Old Men? Long-Term Home Care Utilization in Europe, *Health Economics*, 23(10):1185–1212.

Baranoff, E. G., S. Papadopoulos, and T. W. Sager, 2007, Capital and Risk Revisited: A Structural Equation Model Approach for Life Insurers, *Journal of Risk and Insurance*, 74(3):653–681.

Barczyk, D. and M. Kredler, 2018, The Role of Informal and Formal Care, *Working paper, McGill University, Universidad Carlos III de Madrid.*

Beran, T. N. and C. Violato, 2010, Structural Equation Modeling in Medical Research: A Primer, *BMC Research Notes*, 3(1):267.

Bolin, K., B. Lindgren, and P. Lundborg, 2008, Informal and Formal Care among Single-Living Elderly in Europe, *Health Economics*, 17(3):393–409.

Bonsang, E., 2009, Does Informal Care from Children to Their Elderly Parents Substitute for Formal Care in Europe?, *Journal of Health Economics*, 28(1):143–154.

Brainin, M., J. Tuomilehto, W.-D. Heiss, N. M. Bornstein, P. M. W. Bath, Y. Teuschl, E. Richard, A. Guekht, and T. Quinn, 2015, Post-Stroke Cognitive Decline: an Update and Perspectives for Clinical Research, *European Journal of Neurology*, 22(2):229–316.

Chappell, N. L., 1985, Social Support and the Receipt of Home Care Services, *The Gerontologist*, 25(1):47–54.

Charles, K. K. and P. Sevak, 2005, Can Family Caregiving Substitute for Nursing Home Care?, *Journal of Health Economics*, 24(6):1174–1190.

Christianson, J. B., 1988, The Evaluation of the National Long Term Care Demonstration. 6. The Effect of Channeling on Informal Caregiving, *Health Services Research*, 23(1):99–117.

Coe, N. B. and C. H. Van Houtven, 2009, Caring for Mom and Neglecting Yourself? The Health Effects of Caring for an Elderly Parent, *Health Economics*, 18(9):991–1010.

De Laet, C., J. A. Kanis, A. Odén, H. Johanson, O. Johnell, P. Delmas, J. A. Eisman, H. Kroger, S. Fujiwara, P. Garnero, E. V. McCloskey, D. Mellstrom, L. J. Melton, P. J. Meunier, H. A. Pols, J. Reeve, A. Silman, and A. Tenenhouse, 2005, Body Mass Index as a Predictor of Fracture Risk: A Meta-Analysis, *Osteoporosis International*, 16(11):1330–1338.

De Meijer, C., M. Koopmanschap, T. B. D'Uva, and E. van Doorslaer, 2011, Determinants of Long-Term Care Spending: Age, Time to Death or Disability?, *Journal of Health Economics*, 30(2):425–438.

Denton, M., 1997, The Linkages Between Informal and Formal Care of the Elderly, *Canadian Journal on Aging / La Revue Canadienne du Vieillissement*, 16(1):30–50.

Eling, M. and S. D. Marek, 2013, Corporate Governance and Risk Taking: Evidence From the U.K. and German Insurance Markets, *Journal of Risk and Insurance*, 81(3):653–682.

Finlayson, M., 2002, Changes Predicting Long-Term Care Use Among the Oldest-Old, *The Gerontologist*, 42(4):443–453.

Freedman, V. A. and L. G. Martin, 1999, The Role of Education in Explaining and Forecasting Trends in Functional Limitations among Older Americans, *Demography*, 36(4):461–473.

Fried, L. P., C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, and M. A. McBurnie, 2001, Frailty in Older Adults: Evidence for a Phenotype, *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 56(3):M146–M157.

Fuino, M., I. Rudnytskyi, and J. Wagner, 2019, On the Characteristics of Reporting ADL Limitations and Formal LTC Usage across Europe, *Working Paper, University of Lausanne*.

Fuino, M. and J. Wagner, 2018, Old-Age Care Prevalence in Switzerland: Drivers and Future Development, *European Actuarial Journal*, 8(2):321–362.

Fuino, M. and J. Wagner, 2019, Duration of Long-Term Care: Socio-Economic Drivers, Evolution and Type of Care Interactions, *Working Paper, University of Lausanne.*

Gentili, E., G. Masiero, and F. Mazzonna, 2017, The Role of Culture in Long-Term Care Arrangement Decisions, *Journal of Economic Behavior & Organization*, 143:186–200.

Gregg, E. W., Y. J. Cheng, B. L. Cadwell, G. Imperatore, D. E. Williams, K. M. Flegal, K. M. Narayan, and D. F. Williamson, 2005, Secular Trends in Cardiovascular Disease Risk Factors According to Body Mass Index in US Adults, *The Journal of the American Medical Association*, 293(15):1868–1875.

Hanaoka, C. and E. C. Norton, 2008, Informal and Formal Care for Elderly Persons: How Adult Children's Characteristics Affect the Use of Formal Care in Japan, *Social Science & Medicine*, 67(6):1002–1008.

Headen Jr., A. E., 1993, Economic Disability and Health Determinants of the Hazard of Nursing Home Entry, *The Journal of Human Resources*, 28(1):80–110.

Hooper, D., J. Coughlan, and M. Mullen, 2008, Structural Equation Modelling: Guidelines for Determining Model Fit, *Electronic Journal on Business Research Methods*, 6(1):53–60.

House, J. S., R. C. Kessler, and A. R. Herzog, 1990, Age, Socioeconomic Status, and Health, *The Milbank Quarterly*, 68(3):383–411.

Huisman, M., A. E. Kunst, and J. P. Mackenbach, 2003, Socioeconomic Inequalities in Morbidity among the Elderly; A European Overview, *Social Science & Medicine*, 57(5):861–873.

Jette, A. M., S. Tennstedt, and S. Crawford, 1995, How Does Formal and Informal Community Care Affect Nursing Home Use?, *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 50(1):S4–S12.

Katz, S., A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe, 1963, Studies of Illness in the Aged. The Index of ADL: a Standardized Measure of Biologic and Psychologic Function, *The Journal of the American Medical Association*, 185:94–99.

Kemper, P., 1992, The Use of Formal and Informal Home Care by the Disabled Elderly, *Health Services Research*, 27(4):421–451.

Kim, H. B. and W. Lim, 2015, Long-Term Care Insurance, Informal Care, and Medical Expenditures, *Journal of Public Economics*, 125:128–142.

Kurth, T., J. Gaziano, K. Berger, C. Kase, K. Rexrode, N. Cook, J. Buring, and J. Manson, 2002, Body Mass Index and the Risk of Stroke in Men, *Archives of Internal Medicine*, 162(22):2557–2562.

Langa, K. M., M. E. Chernew, M. U. Kabeto, A. R. Herzog, M. B. Ofstedal, R. J. Willis, R. B. Wallace, L. M. Mucha, W. L. Straus, and A. M. Fendrick, 2001a, National Estimates of the Quantity and Cost of Informal Caregiving for the Elderly with Dementia, *Journal of General Internal Medicine*, 16(11):770–778.

Langa, K. M., M. E. Chernew, M. U. Kabeto, and S. J. Katz, 2001b, The Explosion in Paid Home Health Care in the 1990s : Who Received the Additional Services?, *Medical Care*, 39(2):147–157.

Lo Sasso, A. T. and R. W. Johnson, 2002, Does Informal Care from Adult Children Reduce Nursing Home Admissions for the Elderly?, *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 39(3):279–297.

Lu, B., X. Y. Song, and X. D. Li, 2012, Bayesian Analysis of Multi-Group Nonlinear Structural Equation Models with Application to Behavioral Finance, *Quantitative Finance*, 12(3):477–488.

Lutz, W., W. Sanderson, and S. Scherbov, 2008, The Coming Acceleration of Global Population Ageing, *Nature*, 451(7179):716–719.

Michaud, P.-C. and A. van Soest, 2008, Health and Wealth of Elderly Couples: Causality Tests Using Dynamic Panel Data Models, *Journal of Health Economics*, 27(5):1312–1325.

Mommaerts, C., 2018, Are Coresidence and Nursing Homes Substitutes? Evidence from Medicaid Spend-Down Provisions, *Journal of Health Economics*, 59:125–138.

Muthén, B. O., 1993, Goodness of Fit With Categorical and Other Nonnormal Variables, In K. A. Bollen and J. S. Long, editors, *Testing Structural Equation Models*, chapter 9, pages 205–234. Sage, Newbury Park.

Nachtigall, C., U. Kroehne, F. Funke, and R. Steyer, 2003, (Why) Should We Use SEM? Pros and Cons of Structural Equation Modeling, *Methods Psychological Research Online*, 8(2):1–22.

Newman, S. J., R. Struyk, P. Wright, and M. Rice, 1990, Overwhelming Odds: Caregiving and the Risk of Institutionalization, *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 45(5):S173–S183.

Pezzin, L. E., P. Kemper, and J. Reschovsky, 1996, Does Publicly Provided Home Care Substitute for Family Care? Experimental Evidence with Endogenous Living Arrangements, *The Journal of Human Resources*, 31(3):650–676.

Pi-Sunyer, X., 1991, Health Implications of Obesity, *American Journal of Clinical Nutrition*, 53(6):1595S–1603S.

Rahmqvist, M., 2001, Patient Satisfaction in Relation to Age, Health Status and Other Background Factors: a Model for Comparisons of Care Units, *International Journal for Quality in Health Care*, 13(5):385–390.

Renehan, A. G., M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen, 2008, Body-Mass Index and Incidence of Cancer: a Systematic Review and Meta-Analysis of Prospective Observational Studies, *The Lancet*, 371(9612):569–578.

Rexrode, K., C. Hennekens, W. Willett, G. Colditz, M. Stampfer, J. Rich-Edwards, F. Speizer, and J. Manson, 1997, A Prospective Study of Body Mass Index, Weight Change, and Risk of Stroke in Women, *The Journal of the American Medical Association*, 277(19):1539–1545.

Rodrigues, R., S. Ilinca, and A. E. Schmidt, 2018, Income-Rich and Wealth-Poor? The Impact of Measures of Socio-Economic Status in the Analysis of the Distribution of Long-Term Care Use among Older People, *Health economics*, 27(3):637–646.

Rosseel, Y., 2012, lavaan: An R Package for Structural Equation Modeling, *Journal Of Statistical Software*, 48(2):1–36.

Scheil-Adlung, X., 2015, Long-Term Care Protection for Older Persons: A Review of Coverage Deficits in 46 Countries, *Working paper, International Labour Office*.

Shapiro, E. and R. B. Tate, 1985, Predictors of Long Term Care Facility Use Among the Elderly, *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement*, 4(1):11–19.

Spector, W. D., S. Katz, J. B. Murphy, and J. P. Fulton, 1987, The Hierarchical Relationship Between Activities of Daily Living and Instrumental Activities of Daily Living, *Journal of chronic diseases*, 40(6):481–489.

Swiss Federal Statistical Office, 2019, *L'Enquête Suisse sur la Santé 2017 en Bref. Conception, Méthode, Réalisation*.

Tomiak, M., J.-M. Berthelot, E. Guimond, and C. A. Mustard, 2000, Factors Associated with Nursing-Home Entry for Elders in Manitoba, Canada, *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 55(5):M279–M287.

Torbica, A., S. Calciolari, and G. Fattore, 2015, Does Informal Care Impact Utilization of Healthcare Services? Evidence from a Longitudinal Study of Stroke Patients, *Social Science & Medicine*, 124:29–38.

Van Houtven, C. H. and E. C. Norton, 2004, Informal Care and Health Care Use of Older Adults, *Journal of Health Economics*, 23(6):1159–1180.

Wagner, M. and M. Brandt, 2018, Long-term Care Provision and the Well-Being of Spousal Caregivers: An Analysis of 138 European Regions, *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 73(4):e24–e34.

Weaver, F. M. and B. A. Weaver, 2014, Does Availability of Informal Care within the Household Impact Hospitalisation?, *Health Economics, Policy and Law*, 9(1):71–93.

Witt, B. J., S. J. Jacobsen, S. A. Weston, J. M. Killian, R. A. Meverden, T. G. Allison, G. S. Reeder, and V. L. Roger, 2004, Cardiac Rehabilitation After Myocardial Infarction in the Community, *Journal of the American College of Cardiology*, 44(5):988–996.

Wong, A., R. Elderkamp-De Groot, J. Polder, and J. Van Exel, 2010, Predictors of long-term care utilization by Dutch hospital patients aged 65+, *BMC Health Services Research*, 10:1–14.

Wu, S., R. Wang, Y. Zhao, X. Ma, M. Wu, X. Yan, and J. He, 2013, The Relationship Between Self-Rated Health and Objective Health Status: a Population-Based Study, *BMC Public Health*, 13(1):320.

Chapter 4

The Impact of Catastrophe Events on Insurance Companies' Market Valuation: an Event Study Analysis

This research applies an event study methodology to examine the impact of a selected number of major catastrophe events on the stock valuation of the 87 largest listed non-life insurance companies worldwide. We aim to study if the responses from insurance companies' stock prices depend on the events' nature. In our analysis, we consider airline crashes, earthquakes, hurricanes and winter storms. For each event, we study the significance of its impact on a set of insurers. Depending on the type of the event, the produced effect, its significance and its extent are different. Furthermore, we analyze the relation between the caused valuation effects and the companies' characteristics such as the market capitalization, the revenue, the relevant sectors of insurance, the geographical origin and the split of revenues. Thereby, we try to explain which companies' characteristics drive the stock market response. The obtained results are mixed: for some events, our research gives indications for the management of insurance companies during catastrophe events while in other cases it shows the limits of applicability of event study analyses.

Note: This is joint work with M. Kreutzer (Department of Management & Economics, EBS University) and J. Wagner. The authors are thankful for the comments on earlier versions of this manuscript from participants in the following conferences: Jahrestagung Deutscher Verein für Versicherungswissenschaft (Vienna, Mar. 2016), R in Insurance (London, June 2016), satRday #1 Conference (Budapest, Sept. 2016) and 3rd European Actuarial Journal Conference (Lyon, Sept. 2016). A comprehensive electronic appendix including detailed results is available from the authors upon request.

4.1 Introduction

Major catastrophes typically result in a vast number of claims of significant sizes, which lead to enormous direct losses for insurance companies.[1] Absorbing such losses causes a capital outflow, and thus a decrease in stock's dividends on a short-time horizon. At the same time, the demand of insurance coverage increases in the mid-term after catastrophes. Thereby, it induces increases in premiums and in firms' profits in the longer-term prospective. In this research, we examine the impact and significance of selected natural and man-made catastrophe shocks on the stock valuation of insurance companies. We intend to identify a pattern in the responses of stock prices. Furthermore, we analyze the relationship between firms' characteristics and their stocks' response depending on the type of event.

For studying the influence of an event on a firm's market valuation the event study methodology is the traditional method of choice (see, e.g., Kothari and Warner, 2007). The event study utilizes daily stock prices, which is the most frequently refreshed market evaluation of a company. In an efficient market the stock prices immediately reflect the most recent information updates, which makes event study tests most accurate. The results from statistical tests provide an assessment of the significance of the stock responses. We compare the performance of the six most commonly used parametric tests, the results of which are complemented by six nonparametric test statistics. The parametric tests include classical Student's t-test and tests proposed by Brown and Warner (1980), Brown and Warner (1985), Patell (1976), Boehmer et al. (1991) and Lamb (1995). The nonparametric tests include the sign test, generalized sign test (Cowan, 1992), sign test proposed by Corrado and Zivney (1992), rank test (Corrado, 1989), modified rank test (Corrado and Zivney, 1992) and Wilcoxon signed-rank test.

We consider a sample of the 13 most expensive events in terms of insured losses (cf. Section 4.3.1). The sample includes hurricanes (Katrina, Rita, Wilma, Ike, Irene and Sandy), earthquakes (in Chile, New Zealand and Japan), European winter storms (Kyrill and Klaus), and recent airline crashes (Malaysia Airlines Flight 17 and Germanwings Flight 9525). Market responses on insurance stocks from hurricanes are better studied when compared to the other event types. However, most existing studies consider hurricanes (or other events) individually, using different sets of companies, which makes a comparison impossible. In contrast, our contribution aims to investigate all events consistently within the same framework allowing for comparisons. Gangopadhyay et al. (2010) find strong negative responses of property-liability US companies to Hurricane Katrina and mixed effects to Hurricane Rita. Angbazo and Narayanan (1996) report the negative impact of Hurricane Andrew on property-liability US firms. The results of Lamb (1998) and Lamb (1995) about the effect of Hurricane Andrew coincide with Angbazo and Narayanan (1996) and show that US P&C insurers are unaffected by Hurricane Hugo. Ewing et al. (2006) investigates the development of the market reaction to Hurricane Andrew and Hurricane Floyd and indicates overall negative effect of both hurricanes. Shelor et al. (1992) inspects the reaction of local companies' valuation to the 1989 California Earthquake (positive response) and Marlett et al. (2000) report significantly positive stock reaction, while Yamori and Kobayashi (2002) and Takao et al. (2013) find negative responses of Japanese insurance companies stocks to 1995 Hanshin-Awaji and 2011 Tōhoku earthquakes, respectively. In addition, we test our methodology on validity for the case of 9/11 terrorist attacks, which is also the subject of studies of Chen et al. (2008), Cummins and

[1] `http://institute.swissre.com/research/overview/sigma/`

Lewis (2003), Doherty et al. (2003), Wang and Corbett (2008), Yanase and Yasuda (2010) and Marlett et al. (2003), who report extremely negative effects. A limited number of papers (e.g. Gangopadhyay et al., 2010 and Cummins and Lewis, 2003) use cross-sectional regression analysis to define relevant characteristics of firms for individual events.

Our aim is to study the impact of the selected catastrophe events on 87 listed insurance companies applying the same methodological framework, identifying the presence or absence of a pattern in the market reaction, and determining the companies' characteristics that drive the stock response. After giving an overview of the event study approach and introducing the test statistics, we accomplish preliminary validation of the methodology by approving the performance of the significance measures (parametric and nonparametric tests) for the case of 9/11 terrorist attacks and 31 European non-life insurance companies. For further analysis, we concentrate on one test, the one introduced by Brown and Warner (1985). Then, we accumulate the abnormal returns over the different time periods to show their significance. Furthermore, we split the set of companies into groups according to their market capitalization and business subsector, and outline the difference in reactions. Furthermore, it should be noted that the event study methodology presents some weakness in the shift of the regression coefficients when using a single-Index market model as highlighted by Thomann (2013). In contrast to that work, we show the stability of the results by utilizing several models, namely market-adjusted returns and mean-adjusted returns models. In addition, instead of analyzing the market index, we employ the individual stocks of the companies. Finally, we put the robustness of the results in relation with the estimation window length, as well as the weak sensitivity in relation to market index changes.

All event studies cited above only consider a single event and either the US or the Japanese market. Our research widens the scope by considering several (types of) events and the largest listed companies covering the most important insurance markets including the European market. Fixing the methodological framework and the set of companies for all events allows us to compare the impacts of events, differences of which are mainly due to the events nature, but not to methodological differences. Apart from studies focusing on single events, other papers test various types of events effects using the event study methodology. Chesney et al. (2011) focus on terrorism events impacting on stock, bond and commodity markets. Compared to this paper, as well as to Thomann (2013) we fix the set of the companies, and employ the individual stocks, instead of market index, which gives more precise and accurate insights of the market behavior. We validate our results by several tests, ensuring the robustness to erroneous assumptions. Grace et al. (2014) studies the impact of the sector-specific events on the other market sector contrasting the US banking and insurance industry. Moreover, to the best of our knowledge there are no papers studying the impact of European winter storms on both European and North American insurance markets.

In sum, we derive the following three main findings: (1) No clear pattern in stock responses to catastrophes can be observed. Furthermore, the extent and the responses in terms of valuation (positive or negative) are independent of the type of a catastrophe. The impact is typically mixed, with neither positive nor negative abnormal returns dominating. (2) North American and Western European companies behave differently. North American companies are more influenced by the local events (especially in the case of hurricanes), while Western European insurance groups are sensitive to all events, including major international ones. This can be explained by their broader geographical business exposure. (3) Finally, the height of the market capitalization influences the stock market reaction only in few event cases. Considering the business split of the companies, reinsurance companies are the most sensitive to the catastrophe events, which comes along with their business model.

The remainder of the paper is organized as follows. In Section 4.2 we give a brief overview of the event study methodology by describing formally statistical tests and the regression model on cumulative abnormal returns. In Section 4.3 we derive the set of examined events and describe the selection procedure for the companies and reference indices. In Section 4.4, we present the 9/11 terrorist attacks benchmark case for the validation of our methodology. In Section 4.5 we show the detailed analysis for hurricane, earthquake, winter storm and airline crash events, respectively. In Section 4.6 we summarize and draw our conclusions. A comprehensive Appendix includes the full list of companies, their characteristics and the detailed test statistics for all events.

4.2 Event study methodology and statistical tests

An event study[2] is a statistical toolbox that allows examining the impact of certain events on the stock valuation of a company. Campbell et al. (1997) assume that given the rationality of market participants, the prices of securities immediately incorporate any relevant announcements or information. The idea of the event study is to compare the market valuation of the companies during periods related to an event and other (non-event linked) periods. If the behavior of stocks is significantly different in the event-period, then one concludes that an event produces an impact on the market valuation. Otherwise, one deduces that there is no effect. Thus, an observation period of stock prices is divided into two parts: an estimation window and an event window.[3]

The *estimation window* can be characterized as the period before an event, during which companies experience no particular shocks. The estimation period allows estimating the parameters of the model to see how the securities' prices behave under "normal" conditions. Then, we compare the expected behavior of the securities' prices with observed ones during the event window. Classical event studies[4] make use of an estimation period before the event. There are two other approaches using post-event estimation periods or using the dates around the event window (cf. Henderson, 1990). For example, Wang and Corbett (2008) use post-event choice for September 11, 2001, terrorist attacks. At the same time, Yamori and Kobayashi (2002) report the robustness of results when comparing pre-event and post-event estimation periods in the case of Japanese companies for the 1995 Hanshin-Awaji earthquake. In our study, we follow the classical studies using pre-event estimation periods.

In contrast to the estimation window, the *event window* concentrates on the period around the event. In that period, the market incorporates the information and prices are updated by the

[2]The methodology described in this section has been implemented as an R package available from the authors upon request.

[3]To avoid using the concept of estimation and event windows, alternative approaches can be used. For example, stock returns can be modeled as time series in which events are characterized by an indicator.

[4]See Brown and Warner (1980), Brown and Warner (1985), Lamb (1995), Lamb (1998), Shelor et al. (1992), Angbazo and Narayanan (1996), Ewing et al. (2006), Gangopadhyay et al. (2010), Shelor et al. (1990), Shelor and Cross (1990).

event. The reaction of the market could be delayed (e.g. when the exact magnitude of losses is only known in the days or weeks after the event) or information may be available before the occurrence of the event (e.g. in the case of storms with early warnings), therefore the time frame of the event window can often not be limited to the single day of the event but typically covers several weeks around the event date.

To formally describe both time windows, we first denote by t_e the event date, i.e., the date when the event actually occurred. Then, the set of dates

$$W_\mathrm{event} = \{t_\mathrm{e} - w_\mathrm{b}, t_\mathrm{e} - w_\mathrm{b} + 1, ..., t_\mathrm{e}, ..., t_\mathrm{e} + w_\mathrm{a} - 1, t_\mathrm{e} + w_\mathrm{a}\}, \tag{9}$$

is called the event window, where w_b and w_a parameterize the window counting the days before respectively after the event. The estimation window ends with the day before the beginning of the event window. It is parameterized by its length Δ and defined as follows:

$$W_\mathrm{estim} = \{t_\mathrm{e} - w_\mathrm{b} - \Delta, ..., t_\mathrm{e} - w_\mathrm{b} - 1\}. \tag{10}$$

We illustrate the definitions of the time windows in Figure 4.1.

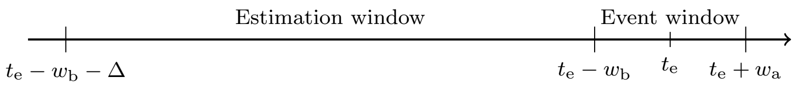

Figure 4.1: Illustration of the estimation and event windows of an event at date t_e.

During the observation period of interest $\mathcal{T} = W_\mathrm{estim} \cup W_\mathrm{event}$, including both the event and the estimation periods, we consider N securities, that are stocks of insurance companies. We call $P_{i,t}$ the price of security i, $i = 1, \ldots, N$, at dates $t \in \mathcal{T}$. To introduce a relative measure and allow for return rate comparisons, we consider the rates of return $R_{i,t}$ instead of the prices $P_{i,t}$ for each security. Finally, Henderson (1990) covers the question of which compounding approach, continuous or periodic, to use. Even though there is no clear evidence that continuous compounding improves normality of returns, most event studies in the literature use this method. Thus, we define the rate of return for security i at the date t as:

$$R_{i,t} = \log\left(\frac{P_{i,t+1}}{P_{i,t}}\right). \tag{11}$$

Furthermore, we assume that each return $R_{i,t}$ can be written as:

$$R_{i,t} = K_{i,t} + \epsilon_{i,t}, \qquad i = 1, \ldots, N, \quad t \in W_\mathrm{estim}, \tag{12}$$

where $K_{i,t}$ is an expected (or predicted) return of security i at time t, given a market model of expected return. $\epsilon_{i,t}$ is an error term (excess return) of security i at time t. For regression models, this term formally refers to residuals, but only for the estimation period.

There is a variety of market models considered in event study papers, such as, for example, the mean-adjusted, the CAPM and the Fama and French 3-factor models (see Fama and French, 1993).[5] Considering all advantages and drawbacks of the above models we use the

[5] Other approaches assume to account for time varying variance and autocorrelation by using, e.g., time series GARCH model. However, as Kolari and Pynnonen, 2011 show, our test statistics are robust to deviations from

single-index market model (SIMM) for our study. The chosen regression model is typically accounted as a valid choice between complexity and performance. Furthermore, it produces a distribution of excess returns that is relatively close to a normal distribution. Nevertheless, this model is the simplest from all regression models and Henderson (1990) states that it performs as good as other more sophisticated models. In addition, we review two other classical event study models, the mean-adjusted-returns and the market-adjusted-returns models. We compare the results with those obtained from our original approach.

First, we introduce the SIMM that can be characterized by the following equation:

$$R_{i,t} = \alpha_i + \beta_i \cdot R_{M,t} + \epsilon_{i,t}, \qquad i = 1, \ldots, N, \quad t \in W_{\text{estim}}, \tag{13}$$

where $R_{M,t}$ is the market proxy return at time t, α_i is a market-independent component of the return of security i, and β_i is a measure of the sensitivity to market changes of security i. The values of a security's return $R_{i,t}$ and of the market return $R_{M,t}$ are the observed ones and can be directly computed from the prices. However, the values of the regression coefficients require estimations. To estimate α_i and β_i we use the ordinary least squares (OLS) method. We expect different return rates during the estimation and the event windows. In order to obtain stable estimates and to avoid shock events in the estimation, the OLS is only performed on the estimation window W_{estim}. Let us denote by $\hat{\alpha_i}$ and $\hat{\beta_i}$ the estimates of α_i and β_i, respectively. These estimates are used for calculating the expected return and the error term $\epsilon_{i,t}$. Patell (1976) emphasizes that for the estimation window W_{estim} these errors can be interpreted as the residuals of the regression. However, for dates in the event window W_{event}, since they are not used in the regression analysis, the interpretation of the deviations will be to consider them as *abnormal returns*. Thus, we define the abnormal return for security i at time $t \in \mathcal{T}$ as the difference between the observed rate of return and the expected one from the regression model. We set:

$$A_{i,t} = R_{i,t} - \hat{\alpha_i} - \hat{\beta_i} \cdot R_{M,t}, \qquad i = 1, \ldots, N, \quad t \in \mathcal{T}. \tag{14}$$

In the simplified mean-adjusted-returns model, the mean of the observed returns $\bar{R}_i$ during the estimation period must be estimated for each security i. This estimate $\bar{R}_i = \sum_{t=t_\text{e}-w_\text{b}-\Delta}^{t_\text{e}-w_\text{b}-1} R_{i,t}$ becomes the expected return, therefore the abnormal return for security i at time t is as follows:

$$A_{i,t} = R_{i,t} - \bar{R}_i, \qquad i = 1, \ldots, N, \quad t \in \mathcal{T}. \tag{15}$$

The market-adjusted-returns model assumes that the expected return at time t equals the return of the market index. Thus, the abnormal return is calculated as follows:

$$A_{i,t} = R_{i,t} - R_{M,t}, \qquad i = 1, \ldots, N, \quad t \in \mathcal{T}. \tag{16}$$

The market-adjusted-returns model can also be interpreted as a special case of the SIMM with $\alpha_i = 0$ and $\beta_i = 1$.
The event study essentially seeks to examine whether the distribution of returns in the event window W_{event} differs significantly from the return distribution in the estimation period W_{estim}. This can be done by analyzing the similarities in the observed and expected returns during the event window W_{event}, which is equivalent to considering the abnormal returns introduced above. We apply both parametric and nonparametric tests to study the significance of differences in

market model assumptions.

the returns.

4.2.1 Parametric tests

The parametric tests assume normal distribution of abnormal returns and in most of the cases focus on the specific characteristic, such as an expected value of cross-sectional abnormal returns. Hence, we introduce a null hypothesis, which is used by all presented parametric tests:

$$H_0 : \mathbb{E}(A_t) = 0, \tag{17}$$

with an alternative one:

$$H_1 : \mathbb{E}(A_t) \neq 0. \tag{18}$$

Each test's statistic is compared with its critical value for significance at levels $\alpha \in \{0.1, 0.05, 0.01\}$, rejecting the null hypothesis when it is in fact true.

All considered tests have so-called Student's shape of the statistics, in other words the statistic is the ratio of the cross-sectional mean and estimated standard deviation. We use the following notation for the estimator of the cross-sectional expected value:

$$\bar{A}_t = \frac{1}{N} \sum_{i=1}^{N} A_{i,t}. \tag{19}$$

Student's t-test. We start from the classical Student's t-test, which assumes cross-sectional independence and normality of cross-sectional returns. This test has the following statistic:

$$\frac{\bar{A}_t}{\hat{S}_{t\text{-test}}}, \tag{20}$$

where

$$\hat{S}_{t\text{-test}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} (A_{i,t} - \bar{A}_t)}. \tag{21}$$

If the null hypothesis H_0 holds, then the statistic has Student's t-distribution with $\Delta - 1$ degrees of freedom. Most often the length of estimation period Δ exceeds 100 days, which allows approximating Student's t-distribution with the standard normal distribution. The main advantage of this test is the fact that it does not require the event-induced variance to be insignificant. However, the requirement of cross-sectional independence extinguishes this useful feature. This test is typically used only as a complement to others due to the limitations from the assumptions.

Tests proposed by Brown and Warner (1980) and Brown and Warner (1985). The authors propose two tests: one with no dependence adjustment and another with crude dependence adjustment. The test with no dependence adjustment was firstly introduced in their 1980-paper and used for monthly data. The test statistic is:

$$\frac{\bar{A}_t}{\hat{S}_{\mathrm{BW1980}}}, \tag{22}$$

where

$$\hat{S}_{\mathrm{BW1980}} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left[\frac{1}{\Delta - 1}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1}\left(A_{i,t} - \bar{A}_i\right)^2\right]}, \tag{23}$$

and

$$\bar{A}_i = \frac{1}{\Delta}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1} A_{i,t}. \tag{24}$$

The second test (from 1985) is tailored for using daily returns and adjusted to cross-sectional dependence, which is crucial for our case. The test statistic is given by:

$$\frac{\bar{A}_t}{\hat{S}_{\mathrm{BW1985}}}, \tag{25}$$

where

$$\hat{S}_{\mathrm{BW1985}} = \sqrt{\frac{1}{\Delta - 1}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1}(\bar{A}_t - \bar{\bar{A}})^2}, \tag{26}$$

and

$$\bar{\bar{A}} = \frac{1}{\Delta}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1} \bar{A}_t. \tag{27}$$

Both statistics have Student's t-distribution with $\Delta - 1$ degrees of freedom, and with the same reasoning as for the traditional t-test, can be approximated by the standard normal distribution.

Test proposed by Patell (1976). This test standardizes the returns before the cross-sectional aggregation. The standardization eliminates the effect of higher standard deviation of returns during the event window. In addition, the standardization does not allow the securities with higher fluctuation to dominate the test. The drawbacks of the test are the same, as for test proposed by Brown and Warner (1980): the test assumes cross-sectional independence and insignificance of event-induced variance. The standardized abnormal return of security i at time $t \in W_{\mathrm{event}}$ is given by:

$$\mathrm{SA}_{i,t} = A_{i,t} \Big/ \hat{s}_i \cdot \sqrt{1 + \frac{1}{\Delta} + \frac{(R_{M,t} - \bar{R}_M)^2}{\sum_{k=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1}(R_{M,k} - \bar{R}_M)^2}}, \tag{28}$$

where

$$\bar{R}_M = \frac{1}{\Delta}\sum_{k=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1} R_{M,t}, \tag{29}$$

and

$$\hat{s}_i = \sqrt{\frac{1}{\Delta - 1}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}-w_\mathrm{b}-1}(A_{i,t} - \bar{A}_i)^2}. \tag{30}$$

Hence, the statistic is given by:

$$\sum_{i=1}^{N}\mathrm{SA}_{i,t} \Big/ \sqrt{N \cdot \frac{\Delta - 2}{\Delta - 4}}. \tag{31}$$

If all assumptions are satisfied, the test statistic is approximately standard normally distributed. For large values of Δ, the fraction $\frac{\Delta-2}{\Delta-4} \approx 1$, and the test statistic can be simplified to:

$$\frac{\sum_{i=1}^{N} \mathrm{SA}_{i,t}}{\sqrt{N}}. \tag{32}$$

Test proposed by Boehmer et al. (1991). This hybrid of classical t-test and test proposed by Patell (1976) reduces the issue of event-induced variance to the minimum, but still assumes cross-sectional independence. As in Patell's test the abnormal returns are standardized before the aggregation, but in contrast use the t-shape of the test, as in the t-test. The Student's t-distribution is assumed, if the null hypothesis H_0 holds. The test statistics is as follows:

$$\frac{\overline{\mathrm{SA}}_t}{\hat{S}_{\mathrm{BMP}}}, \tag{33}$$

where

$$\overline{\mathrm{SA}}_t = \frac{1}{N}\sum_{i=1}^{N} \mathrm{SA}_{i,t}, \tag{34}$$

and

$$\hat{S}_{\mathrm{BMP}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(\mathrm{SA}_{i,t} - \overline{\mathrm{SA}}_t)}. \tag{35}$$

Test proposed by Lamb (1995). The statistic of the last considered parametric test is close to the on from the Brown and Warner (1985) test, and differs only by a correction coefficient:

$$\frac{\bar{A}_t}{\hat{S}_{\mathrm{Lamb}}}, \tag{36}$$

where

$$\hat{S}_{\mathrm{Lamb}} = \sigma \cdot \sqrt{1 + \frac{1}{\Delta} + \frac{(R_{M,t} - \bar{R}_M)^2}{\sum_{k=t_{\mathrm{e}}-w_{\mathrm{b}}-\Delta}^{t_{\mathrm{e}}-w_{\mathrm{b}}-1}(R_{M,k} - \bar{R}_M)^2}}, \tag{37}$$

and

$$\sigma = \sqrt{\frac{1}{\Delta - 1}\sum_{t=t_{\mathrm{e}}-w_{\mathrm{b}}-\Delta}^{t_{\mathrm{e}}-w_{\mathrm{b}}-1}(\bar{A}_i - \bar{\bar{A}})^2}. \tag{38}$$

4.2.2 Nonparametric tests

All previous tests explicitly assume that the abnormal returns $A_{i,t}$ are normally distributed. This strong assumption cannot always be met with real-world historical data. Thus, parametric tests are frequently supported by nonparametric tests in the event study literature.[6]

Sign test. One of such tests is the sign test, a simple binomial test, which allows confirming that a couple of companies are not dominating the result. The test is discussed in Brown and Warner (1980) and Boehmer et al. (1991). It establishes whether the frequency of positive abnormal returns at a given day in the event window is significantly different from 0.5 (i.e., 50%).

[6] Statistics of all considered nonparametric tests, except Wilcoxon signed-rank test, yield approximately standard normal distribution under the null hypothesis. The latter test has no simple expressible distribution.

However, the naive assumption that the number of positive abnormal returns equals to the number of negative ones with absence of event effect does not hold if the distribution of returns is skewed (see, for example, Brown and Warner, 1980). The statistic of the test is calculated as follows:

$$\left| \frac{\sum_{i=1}^{N} \mathbb{1}_{\{A_{i,t}>0\}}}{N} - \frac{1}{2} \right| \cdot 2\sqrt{N}, \tag{39}$$

where $\mathbb{1}_{\{A_{i,t}>0\}}$ is the indicator function taking the value 1 if $A_{i,t} > 0$ and the value 0 otherwise.

Generalized sign test. The generalized version of the sign test compares the frequency of positive abnormal returns in the event window to the frequency of positive abnormal returns in the estimation period. In other words, the test establishes whether the frequencies in the two periods are significantly different, i.e. analyzing a potential asymmetry in abnormal returns. We consider the test statistic definition from Cowan (1992):

$$\left| \frac{\sum_{i=1}^{N} \mathbb{1}_{\{A_{i,t}>0\}}}{N} - \hat{p} \right| \cdot \sqrt{\frac{N}{\hat{p}(1-\hat{p})}}, \tag{40}$$

where

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\Delta} \sum_{t=t_{\mathrm{e}}-w_{\mathrm{b}}-\Delta}^{t_{\mathrm{e}}-w_{\mathrm{b}}-1} \mathbb{1}_{\{A_{i,t}>0\}}. \tag{41}$$

Sign test proposed by Corrado and Zivney (1992). These authors also propose a test, which is specified under the condition of an asymmetrical distribution of abnormal returns. Essentially, the test uses the estimated median for each security i on the whole period $\mathcal{T}$ including both event and estimation windows. Let m_i denote the sample median for security i and consider the respective “sign” $G_{i,t}$ of the abnormal returns defined by

$$G_{i,t} = \mathrm{sign}(A_{i,t} - m_i), \tag{42}$$

where sign is a sign function, specified as follows:

$$\mathrm{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{43}$$

Then, the statistics is

$$\frac{\bar{G}_t}{\sqrt{N} \cdot \hat{S}_{\mathrm{sign}}}, \tag{44}$$

where

$$\hat{S}_{\mathrm{sign}} = \sqrt{\frac{1}{\Delta + w_{\mathrm{b}} + w_{\mathrm{a}} + 1} \sum_{t=t_{\mathrm{e}}-w_{\mathrm{b}}-\Delta}^{t_{\mathrm{e}}+w_{\mathrm{a}}} \left(\frac{1}{N} \sum_{i=1}^{N} G_{i,t} \right)^2}. \tag{45}$$

Rank test. In addition to the sign test, Corrado (1989) covers the test, which involves the ranks of the abnormal returns. This test does not distinguish between the estimation and the event periods, and uses the whole sample period. Using the Corrado’s original notation, we denote with $K_{i,t}$ the rank of abnormal return $A_{i,t}$ at time $t \in \mathcal{T}$. In other words, the rank one is assigned to the smallest value of $A_{i,t}$ and the rank of $\Delta + w_\mathrm{b} + w_\mathrm{a} + 1$ is referred to the largest one. The average rank equals to $\bar{K} = \frac{1}{2} + (\Delta + w_\mathrm{b} + w_\mathrm{a} + 1)/2$ for all securities, which leads to the statistic:

$$\frac{\frac{1}{N}\sum_{i=1}^{N}(K_{i,t} - \bar{K})}{\hat{S}_\mathrm{rank}}, \tag{46}$$

where

$$\hat{S}_\mathrm{rank} = \sqrt{\frac{1}{\Delta + w_\mathrm{b} + w_\mathrm{a} + 1}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}+w_\mathrm{a}}\left(\frac{1}{N}\sum_{i=1}^{N}(K_{i,t} - \bar{K})\right)^2}. \tag{47}$$

Modified rank test. To allow for missing observations in the series of abnormal returns, Corrado and Zivney (1992) propose to modify the above test statistic by standardizing the respective ranks of abnormal returns, i.e.,

$$\mathrm{SK}_{i,t} = \frac{K_{i,t}}{1 + \mathrm{NM}_i}, \tag{48}$$

where NM_i is the number of non-missing abnormal returns for security i in the observed period $\mathcal{T}$. Let the number of non-missing abnormal returns for the day t be denoted by N_t (out of N considered securities), then the modified rank test statistic is as follows:

$$\frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathrm{SK}_{i,t} - \frac{1}{2})}{\hat{S}_\mathrm{mrank}}, \tag{49}$$

with

$$\hat{S}_\mathrm{mrank} = \sqrt{\frac{1}{\Delta + w_\mathrm{b} + w_\mathrm{a} + 1}\sum_{t=t_\mathrm{e}-w_\mathrm{b}-\Delta}^{t_\mathrm{e}+w_\mathrm{a}}\left(\frac{1}{\sqrt{N_t}}\sum_{i=1}^{N_t}(\mathrm{SK}_{i,t} - \frac{1}{2})\right)^2}. \tag{50}$$

Wilcoxon signed-rank test. It is also possible to apply Wilcoxon signed-rank test, as reported by Wilcoxon (1945). The test examines whether the difference between expected and actual returns follows a symmetric distribution around zero. To proceed the test, for each date t in the event window, the sample of $A_{i,t}$, excluding those $A_{i,t}$ which equal to zero, are considered. As a result, a sample of size N_r is obtained. Each non-zero $A_{i,t}$ is substituted by its respective rank $\tilde{K}_{i,t}$ yielding the test statistic:

$$\sum_{i=1}^{N_r}\left[\mathrm{sign}(A_{i,t}) \cdot \tilde{K}_{i,t}\right]. \tag{51}$$

4.2.3 Cumulative abnormal returns testing

Previous tests focus on a single day in the event window W_event. This entails problems related to lifetime of events and their effects. For instance, the lifetime of hurricanes and storms are not limited to one day, while statistics of consequent days could differ by sign. Also, in the case of the 9/11 terrorist attacks, the effect of the event was extended to weeks. To address this issue Lamb (1995) and Angbazo and Narayanan (1996) test in their empirical study not only a single day of event, but the whole event period (or a part of it). This allows obtaining a single measure, which will reflect the overall effect of the event. We define the cross-sectional cumulative abnormal returns (CCAR) from t_1 to t_2 in the event window as a sum of the cross-sectional means of abnormal returns

$$\mathrm{CCAR}(t_1, t_2) = \sum_{t=t_1}^{t_2} \bar{A}_t, \tag{52}$$

which is tested on equality to zero. The statistic is as follows:

$$\frac{\sum_{t=t_1}^{t_2} \frac{\bar{A}_t}{\hat{S}}}{\sqrt{t_2 - t_1 + 1}}, \tag{53}$$

where $\hat{S}$ represents either $\hat{S}_{\mathrm{BW1985}}$ or $\hat{S}_{\mathrm{Lamb}}$ (cf. Equations 26 and 37). The latter standard deviation depends on the time t, thus cannot be taken out from the summation sign. The test statistic has standard normal distribution in approximation, if the null hypothesis holds.

4.2.4 Cross-sectional regression analysis

In an additional step, we analyze the relationship between the market responses and companies' characteristics. To define and determine the relevance of the companies' characteristics, Lamb (1995) and Shelor et al. (1992) use cross-sectional regression analyses, which employ the estimation of regression coefficients and an examination of their significance. For this approach, we study the i^{th} company's cumulative abnormal returns (CAR) as a function of five independent variables: the company's market capitalization, its business subsector, gross premiums written in life, gross premiums written in non-life and the geographical origin of the company. The market capitalization reflects the size of the company, which allows for studying how the stock market reaction depends on a firm's size (the same characteristic is also used by Gangopadhyay et al., 2010). Regulatory restrictions disallow companies to mix non-life and life lines of business and a group holds them in separate entities. However, at the group level and from an investor's point of view, the gross premiums written in life and non-life can represent the diversification of companies. Finally, if we consider companies both from Europe and North America, the geographical origin of the company can indicate the risk exposure to a certain event.

Before estimating regression parameters, the CAR and the independent variables are standardized, i.e. from each value we subtract the cross-sectional mean and then divide by the standard deviation (these variables are marked by a "$\sim$"). Defining the i^{th} company's CAR as

$$\mathrm{CAR}_i(t_1, t_2) = \sum_{t=t_1}^{t_2} A_{i,t}, \tag{54}$$

the model can be described by the following regression equation for company i:

$$\widetilde{\mathrm{CAR}}_i(t_1, t_2) = \gamma_0 + \gamma_1 \cdot \widetilde{\mathrm{MC}}_i + \gamma_2 \cdot \mathrm{SUB}_i + \gamma_3 \cdot \widetilde{\mathrm{GPWL}}_i + \gamma_4 \cdot \widetilde{\mathrm{GPWNL}}_i + \gamma_5 \cdot \mathrm{GEO}_i + \epsilon_i, \tag{55}$$

where:

- $\widetilde{\mathrm{CAR}}_i(t_1, t_2)$ is the standardized CAR of the company,
- $\widetilde{\mathrm{MC}}_i$ is the standardized market capitalization in EUR of the company on December 31 in the year of the event,

- SUB_i is the categorical variable of the company's subsector, with possible values: "Property & Casualty Insurance", "Full Line Insurance" and "Reinsurance",
- $\widetilde{\text{GPWL}}_i$ are the company's standardized gross premiums written in the life line of the business in EUR in the year of the event,
- $\widetilde{\text{GPWNL}}_i$ are the company's standardized gross premiums written in the non-life line of the business in EUR in the year of the event,
- GEO_i is the dummy (binary) variable, which represents the geographical origin of the company. We consider two regions: if the origin of the company is Western Europe, then $\text{GEO}_i = 1$, otherwise if the origin is North America, then $\text{GEO}_i = 0$,
- γ_j for $j = 0, ..., 5$ are the regression coefficients,
- ϵ_i is the error term.

4.3 Database of events, companies and market values

4.3.1 Events and model parameters

Previous works have concentrated on the analysis of one single event, such as hurricanes or earthquakes. In contrast, our research applies the same methodology and framework for several events. This allows to compare the impact of different events of the same type and to compare the consequences from the different types of catastrophes. First, we define the time range, for which such events are collected. Existing papers focus mostly on events before 2005. However, the reaction of market valuation might not be a constant and develop through the years. Thus, we focus on more recent events in a period starting from January 1, 2005 to September 1, 2015. The scope of our paper includes natural and man-made catastrophes, the data of which is well described in annual statistics available from Swiss Re in their *sigma* publication series[7] (cf. Tables "The 20 most costly insurance losses" and "The 20 worst catastrophes in terms of victims" for each given year starting from 2006). It seems natural to assume that the costliest catastrophes in terms of insured losses influence insurance companies most. Furthermore, it may be interesting to test how companies' stocks react to worst catastrophes in terms of victims, which may not lead to direct non-life losses. The list of events we consider, with corresponding dates, insured losses and number of victims is reported in Table 4.1.

The event dates in the Swiss Re *sigma* reports are approximate. A simplest approach would assume to consider for the event dates the dates when the events occurred (or started). In our analysis, we set the event dates to dates, when the market participants get aware that the event happens. In other words, the event date is the date, when the information is already incorporated in the prices of securities. However, it is not clear how to define, for example, the event dates of the hurricanes. Thus, we focus not only on a single date, but on all dates in the event window. The definition of the event window boundaries must also be set with care. Studies typically use a relatively large event window (often more than two trading weeks). In fact, using large event windows might lead to incorporation of other events: for instance, in the case of Hurricane Katrina, on September 5 Mandala Airlines Flight 091 crashed into a crowded residential area in Medan, which is an event itself. Therefore, we try to accurately

[7] `http://institute.swissre.com/research/overview/sigma/`

#	Name	Victims	Insured loss in mUSD	Event start	Event end	Estim. start	Estim. end	Country
1	Hurricane Katrina	1 836	78 638	2005-08-23	2005-09-02	2005-03-07	2005-08-22	USA et al.
2	Hurricane Rita	34	12 240	2005-09-19	2005-09-30	2005-03-07	2005-08-22	USA et al.
3	Hurricane Wilma	35	15 234	2005-10-17	2005-10-28	2005-03-07	2005-08-22	USA et al.
4	Hurricane Ike	136	22 258	2008-09-02	2008-09-12	2008-03-17	2008-09-01	USA et al.
5	Hurricane Irene	50	6 134	2011-08-20	2011-09-02	2011-03-04	2011-08-19	USA et al.
6	Hurricane Sandy	237	36 079	2012-10-22	2012-11-02	2012-05-06	2012-10-21	USA et al.
7	Earthquake in Chile	562	8 682	2010-03-01	2010-03-05	2009-09-13	2010-02-28	Chile
8	Christchurch earthquake	181	16 836	2011-02-22	2011-03-04	2010-09-04	2011-02-19	New Zealand
9	Tōhoku earthquake	18 520	36 828	2011-03-11	2011-03-24	2010-09-04	2011-02-19	Japan
10	Winter storm Kyrill	54	6 959	2007-01-17	2007-01-24	2006-08-01	2007-01-16	Germany, et al.
11	Winter storm Klaus	25	3 501	2009-01-23	2009-01-28	2008-08-07	2009-01-22	France, Spain
12	Malaysia Airlines Flight	298	–	2014-07-17	2014-07-25	2014-01-29	2014-07-16	Ukraine
13	Germanwings Flight	150	–	2015-03-24	2015-04-02	2014-10-06	2015-03-23	France

Table 4.1: Selection of catastrophe events and their characteristics for the event study analysis.

choose the event windows to be short enough not to overlap with other occurrences, while at the same time to be long enough to reflect the influence. The event window for hurricanes is set to two calendar weeks (or approximately ten trading days) to include the whole lifetime of the hurricane (the lifetime lasts typically from one to two calendar weeks). The starting date for such events is considered as the beginning of the hurricane's lifetime. In contrast, the starting date of the earthquakes is the day when the earthquake took place (the event date). We set the event windows to be equal to two calendar weeks (approximately ten trading days) for earthquakes, and for storms and man-made catastrophes we use one calendar week.

For all events the estimation window length Δ is set to 120 trading days (see, e.g., Angbazo and Narayanan, 1996), and for the market index we use STOXX Global 1800 as a reference. Both choices are validated on the example of the 9/11 terrorist attacks case in Section 4.4.

4.3.2 Companies and indices

To select publicly-traded insurance companies, we chose an industrial classification system. There is a variety of classifications, such as the Standard Industrial Classification (SIC) and the Global Industry Classification Standard (GICS) with their respective advantages and drawbacks. On the other hand, an event study is a complex problem, which requires a coherent approach. One important aspect will be the market index to use in the SIMM. To verify the robustness of our approach, we will use different indices provided by STOXX Limited. This index provider uses the Industry Classification Benchmark (ICB) launched by Dow Jones and FTSE in 2005. Thus, and to be consistent, we also use this industrial taxonomy.

The initial list of 665 publicly-traded insurance companies provided by our data provider Bloomberg is selected by identifying those, who have the ICB supersector number code equal to 8500 that corresponds to insurance companies. Then, we build the data base of key annual characteristics, namely market capitalization (`CUR_MKT_CAP`), gross premiums written in non-life business (`IS_GROSS_PREM_WRITTEN_NL`) and gross premiums written in life business (`IS_GROSS_PREM_WRITTEN_LIFE`). The values are obtained in EUR currency.[8] In addition the information concerning the geographical origin (`CNTRY_OF_RISK`) and the sector/subsector of the company (`ICB_SECTOR_NAME/ICB_SUBSECTOR_NAME`) are gathered. This allows grouping the

[8]Each security default currency different from EUR is converted to EUR at the exchange rate of the time of the measurement.

initial list of companies by sector (non-life, life) and subsector: Full Line Insurance (FL), Insurance Brokers, Property & Casualty Insurance (P&C), Reinsurance (Re) and Life Insurance. The companies in the Insurance Brokers group and in the Life Insurance group are excluded from the list. In Life Insurance, natural catastrophes are unlikely to produce direct extreme losses, and, moreover, the risk of life insurance is typically diversified across different geographic regions. Insurance brokers have no direct losses from catastrophes, and typically the performance is measured annually, thus it is very unlikely for insurance brokers to be directly influenced. The default geographical classification provided by Bloomberg is used to split the companies by region. Apart from Western Europe and North America the number of publicly-traded insurance companies of other regions is small. Thus, we consider only the two mentioned regions.

For each company from the reduced list of 177 companies we have gathered historical daily "Close" stock prices for the time range between January 1, 2005 and September 1, 2015.[9] The "Close" prices are selected due to the fact, that market indices are usually calculated on that basis. These prices also include the event information from the given day (while "Open" prices would not, see Figure 4.2 for an illustration with an event registered at day 2).

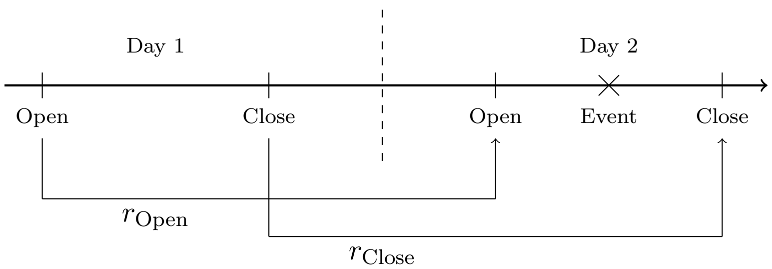

Figure 4.2: Rates of return for "Open" and "Close" prices. The "Close" price on day 2 contains information about the event that occurred on day 2.

To have the same set of companies for each event, companies with enough data for all events should be selected. Thus, for each company the percentage of non-missing observation during the observation period $\mathcal{T}$ of each event is computed, and then, we select those companies, which have more than 60% of entries for all events (accounting for weekends, holidays and so on). Using more than 60% of observation allows us to use more dates in the estimation period, therefore to obtain more stability in market model parameters' estimates. For the selected events in our study and their respective time windows, only 87 companies remain with enough data. In addition, and to show if the market acts differently on companies according to their subsector of the business, geographical origin and their sizes (here: the market capitalization), the companies are grouped by these characteristics. Initially, the companies are split into the North America and Western Europe groups, according to Bloomberg regional split. Then, instead of using the market capitalization of the year prior the event, we compute for each company the average of the annual market capitalization (only for years before the events). This allows having the same groups of companies for all events (otherwise, companies could be included to different groups for different events, which should be avoided). Finally, separately for each geographical group we estimate the quantiles of the average market capitalization, which are used to define groups of small-, middle- and large-sized companies. The larger companies are those, for which the market capitalization is above the 66.7%-quantile value, while middle ones have the market

[9] For setting up our reference case of September 11 terrorist attacks the data set of 2001 is also compiled.

capitalization between the 33.3%- and 66.7%-quantiles. Accordingly, the small companies are defined as companies with an average market capitalization less than 33.3%-quantile. The obtained quantiles are close enough to values defined by taxonomies' definitions. Independently from their size, companies are also grouped by their subsector of the business. Table 4.2 shows the number of companies in each group. The summary of the companies with their respective characteristics can be seen in the appendix.

	North America				Western Europe			
	P&C	FL	Re	Total	P&C	FL	Re	Total
Small	17	1	0	18	5	6	0	11
Middle	12	2	4	18	4	6	1	11
Large	12	3	3	18	3	5	3	11
Total	41	6	7	54	12	17	4	33

Table 4.2: Numbers of companies in each group.

4.4 Reference case: event study of 9/11 terrorist attacks

The event study approach and methodology have strong assumptions. In order to investigate the performance of the method we examine the commonly accepted (*a priori*) significant event of the 2001 9/11 terrorist attacks, which is, in addition, well-studied in previous works (see Chen et al. (2008), Cummins and Lewis (2003), Doherty et al. (2003), Wang and Corbett (2008) and Yanase and Yasuda (2010)).

Thomann (2013) reports that the classical event study results may be biased due to the increase in both volatility and correlation. This paper also discusses the positive shift in the relationship of the individual stock's returns and the market's returns (so-called *beta*), which may doubt the validity of the event study. However, the paper instead of using individual stocks with further cross-sectional averaging employs an equally-weighted index of P&C stocks. At the same time, this index does not consist of the fixed list of companies, but varies from year to year, which may affect the result. Finally, mixing European and North American companies also could shift the result (further, we show that for some events stock prices behave differently depending on the geographical split). To address the problem of increased volatility during the events we support our findings by the test proposed by Boehmer et al. (1991) (their test incorporates event induced variance). To prove the event study validity and robustness we complement the classical SIMM by using a mean-adjusted returns model, which is not influenced by increases in *beta*.

As a direct consequence of 9/11, all main exchanges stopped trading because of the risk of follow-up terrorist attacks. Moreover, NYSE remained closed until the following Monday 9/17. In our context, this means that there is a lack of data for US companies, implying that for this event we can only consider non-US traded companies, particularly Western European insurers (cf. the firms sample discussed in the previous section). The 50%- and the 75%-quantiles of the market capitalization of these companies are respectively 1 419 mEUR and 6 015 mEUR, according to which the companies are split into the large-, middle- and small-sized company groups. The table of the companies with their characteristics is reported in the appendix, while the numbers of companies in each group is presented in Table 4.3.

	FL	P&C	Re	Total
Small	9	7	0	16
Mid	2	3	2	7
Large	6	0	2	8
Total	17	10	4	31

Table 4.3: Numbers of companies in each group used for the 9/11 terrorist attacks event case.

First, we need to determine the values of the parameters which we will base on choices made in previous studies. Then, we apply our methodology to the data to derive the results. Finally, to validate the sensitivity and stability of the results, we vary the parameters' values in a sensitivity analysis and compare the output obtained from the different tests.

For this analysis, we focus on the SIMM, for which the STOXX Global 1800 is used as the market proxy. Since this index represents the behavior of the global world economy, we expect higher levels of deviation of abnormal returns than in the case of more specific industry indices (like insurance indices). The length of the estimation period is set to $\Delta = 120$, which stems from the value used in Angbazo and Narayanan (1996). This period is considered long enough to give reliable estimates of the market model parameters and at the same time avoids incorporating any other events. In our case, the estimation period of 9/11 terrorist attacks given the value of Δ does not overlap with any event windows from other major catastrophes.

It is natural to assume that there has been no public information about these terrorist attacks available before September 11, therefore we set $w_{\mathrm{b}} = 0$. Wang and Corbett (2008) cover days in the event window until October 12 due to White House proposal for a new reinsurance program. This proposal was released on October 15, and is considered as an event itself. In our first approach, we have also taken this date for the limit of the event window. However, we have observed that for the European stock insurance market the last two trading weeks (early October) do not show any significance. Hence, we set the last date in the event window to September 28 (thus we have $w_{\mathrm{a}} = 17$). Using only European companies enables us to also include the dates between September 11 and September 17. None of the considered companies have missing entries during the event window, therefore the percentage of the companies used in our analysis is 100%.

Parametric tests. Table 4.4 shows the results of the parametric tests. We start from the test proposed by Patell (1976), which allows for the event-induced variance. This test produces the highest number of significant statistics and seems to reject the null hypothesis too frequently. The t-test and hybrid test (BMP) proposed by Boehmer et al. (1991) produce the smallest number of significant items and show similar values for the statistics. However, both tests do not consider the cross-sectional dependence which cannot be neglected in our context. The same applies to the BW1980 test proposed by Brown and Warner (1980). The remaining two tests BW1985 and Lamb (Brown and Warner (1985) and Lamb (1995)) theoretically incorporate cross-sectional correlation and have similar statistics. Therefore, we consider these two tests as a benchmark. In our further analysis, we will give preference to the BW1985 test introduced by Brown and Warner (1985).

Date	W.day	$\bar{A}_t$, %	BW1980		BW1985		t-test		Patell		BMP		Lamb	
09-11	Tues	-5.624	-15.090	***	-11.229	***	-4.961	***	-19.424	***	-4.855	***	-10.942	***
09-12	Wed	-3.664	-9.829	***	-7.314	***	-2.674	**	-11.190	***	-1.934	*	-7.280	***
09-13	Thurs	-0.286	-0.767		-0.571		-0.282		-1.344		-0.353		-0.568	
09-14	Fri	-3.097	-8.310	***	-6.184	***	-4.081	***	-10.881	***	-4.061	***	-6.111	***
09-17	Mon	0.673	1.806	*	1.344		0.763		2.767	***	0.993		1.302	
09-18	Tues	-0.512	-1.373		-1.022		-0.722		-1.266		-0.487		-1.016	
09-19	Wed	-1.061	-2.846	***	-2.118	**	-1.022		-1.560		-0.603		-2.101	**
09-20	Thurs	-5.064	-13.587	***	-10.111	***	-5.516	***	-15.608	***	-4.686	***	-9.814	***
09-21	Fri	-4.292	-11.515	***	-8.568	***	-4.401	***	-16.331	***	-4.829	***	-8.505	***
09-24	Mon	3.496	9.381	***	6.981	***	3.418	***	10.527	***	3.491	***	6.750	***
09-25	Tues	1.573	4.221	***	3.141	***	1.820	*	4.479	***	1.531		3.126	***
09-26	Wed	2.475	6.641	***	4.942	***	3.691	***	8.779	***	3.361	***	4.921	***
09-27	Thurs	0.400	1.074		0.799		0.396		1.905	*	0.525		0.792	
09-28	Fri	1.437	3.855	***	2.869	***	1.918	*	6.270	***	2.211	**	2.788	***

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.4: Parametric tests statistics and significance for the 9/11 terrorist attacks.

Nonparametric tests. Table 4.5 presents the output of the nonparametric tests. The dates where we observe significant results are the same for all tests, except September 25 and September 28. On September 25 the sign, generalized sign (G.sign) and Wilcoxon signed-rank (Wlcx) tests produce different results. This date is also a subject to differing results from parametric tests. Most likely, it can be explained by the fact that assumptions of normality (for parametric tests) and symmetry of distribution (for sign, generalized sign and Wilcoxon tests) are not satisfied. Furthermore, on September 28 only the Wilcoxon signed-rank test shows significance. For this case, ignoring dependences plays a crucial role in Wilcoxon test misspecification. Nevertheless, there is no perfect match between significant dates of parametric and nonparametric tests; the significant dates found through nonparametric tests correspond to the dates with extreme values of the parametric statistics. In the sequel, we will skip using nonparametric tests, and we will use them only where required (e.g. when the parametric and nonparametric tests would yield very different results).

Date	W.day	Sign		G.sign		C.sign		Rank		M.rank		Wlcx	
09-11	Tues	-3.413	***	-3.628	***	-2.019	**	-2.828	***	-2.907	***	48.000	***
09-12	Wed	-3.413	***	-3.628	***	-2.131	**	-2.242	**	-2.331	**	98.000	***
09-13	Thurs	-0.180		-0.392		0.336		-0.187		-0.199		240.000	
09-14	Fri	-3.413	***	-3.628	***	-2.131	**	-2.693	***	-2.789	***	52.000	***
09-17	Mon	0.180		-0.033		0.112		0.134		0.110		268.000	
09-18	Tues	-0.539		-0.752		-0.112		-0.568		-0.579		207.000	
09-19	Wed	-0.898		-1.111		-0.561		-0.510		-0.532		207.000	
09-20	Thurs	-3.413	***	-3.628	***	-2.131	**	-3.054	***	-3.152	***	33.000	***
09-21	Fri	-3.413	***	-3.628	***	-1.906	*	-2.873	***	-2.943	***	57.000	***
09-24	Mon	3.053	***	2.843	***	1.906	*	2.537	**	2.611	***	403.000	***
09-25	Tues	1.976	**	1.764	*	1.234		1.407		1.446		348.000	**
09-26	Wed	3.772	***	3.562	***	1.906	*	2.570	**	2.642	***	430.000	***
09-27	Thurs	-0.180		-0.392		-0.336		-0.071		-0.099		253.000	
09-28	Fri	1.616		1.405		1.009		1.244		1.286		347.000	**

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.5: Nonparametric tests statistics and significance for the 9/11 terrorist attacks.

For the whole set of European companies, we observe negative signs in the statistics for dates in the first two weeks after the event (except on September 17). The positive statistics on September 17 may be related to three different but not mutually exclusive causes. The first and most abstract reason is random occurrence. Also, we should consider that the U.S. stock market has reopened exactly on September 17, which could have positively influenced the investors' views. On the other hand, the statistic of this day does not show any significance in the nonparametric tests, which can be a signal, that there are a few companies with extreme abnormal returns driving the test result. This reason is also confirmed by splitting the set of companies in groups according to their business lines. We observe different signs for the statistics, e.g., for full-line insurers and reinsurers (positive sign) and for P&C companies (negative sign), see below.

In contrast to the first two weeks, the third trading week shows only positive statistics. The highest absolute values of the statistics occur on September 11 and September 20, both in the pattern of the first two trading weeks with negative signs. All days except September 13, 18 and 27 show significant statistics at least at level 0.1. Accordingly, the impact develops through the first three weeks after the event date, also probably due to new information releases.

The negative abnormal returns generated during the first two weeks coincide with the initial intuition: companies must absorb the losses and will have to pay claims (of not yet fully known amount) to policyholders, which leads to releasing the capital. That is why within the first two weeks the effect is negative. At the same time, the positive statistics in the third week are also consistent with the rational explanation "gaining from loss". The 9/11 terrorist attacks dramatically have changed insurer's beliefs about the risk of such attacks. After this unexpected event, most insurers either have restricted the coverage or significantly increased the premiums. Taking into account an increased demand, this leads to future potential increases in the profits on the long run. Furthermore, after natural catastrophes the regulatory institutions typically ease the requirements to insurance companies, which allows firms to rise the premiums. Cummins and Lewis (2003) outline such factors as the absence of follow-up attacks, reconsideration of loss estimates, etc. However, at this point it is hardly possible to distinguish which factor is the major one. Our results are consistent with previous studies showing only minor differences. These variations are likely due to the market distinction (Europe) and the event window (starting earlier).

Cumulative abnormal returns testing. To approve and validate our initial conclusion we also test the cross-sectional cumulative abnormal returns CCAR for different time ranges $[t_1, t_2]$. Results are reported in Table 4.6. Both test statistics proposed by Lamb (1995) and Brown and Warner (1985) show similar results. For all time ranges the statistics are significant at the 0.01 level, except for the time range between September 17 and September 28, which refers to the second and third weeks after the event. The absence of significance can be explained by the fact, that negative statistics of the second week are compensated by the positive sign of the third week (see separate periods September 17-21 and September 24-28). These results are also in line with our daily analysis.

Cross-sectional regression analysis. To examine the relevance of companies' characteristics we also conduct a regression analysis. We run the regression for each single day in the event window, as well as for all possible time periods inside the event study. Unfortunately, the result is unstable to make any substantial conclusion. Almost for all time periods the estimated

t_1	t_2	Lamb		BW1985	
09-11	09-12	-12.885	***	-13.112	***
09-11	09-13	-10.849	***	-11.035	***
09-11	09-14	-12.451	***	-12.649	***
09-11	09-21	-15.012	***	-15.258	***
09-11	09-28	-7.125	***	-7.227	***
09-17	09-21	-9.004	***	-9.157	***
09-17	09-28	-0.556		-0.551	
09-24	09-28	8.218	***	8.377	***

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.6: Values of the tests of CCAR for the 9/11 terrorist attacks event case.

parameters are insignificant, except for September 11 and for September 17. These two dates are remarkable, because on the first one the actual event has occurred, and on the second one the U.S. stock exchanges were reopened, which might also be considered as the event date for U.S. companies. The respective parameters are presented in Table 4.7. From first sight the result might look contradictory: the signs of the estimated parameters are perfectly the opposite. However, considering the different signs of the statistics on September 11 and September 17, it can be explained. The positive sign of the parameter estimate for market capitalization means that the larger the company the smaller the impact. In other words, large companies are more stable and less exposed to the effect of the event. For the September 17, the conclusion from negative signs of statistics is the same: the larger company, the lower the impact. The sign of the reinsurance modality is negative for September 11 (respectively for September 17 positive), which implies that the reinsurance companies affected much stronger than the other types. This conclusion, that large companies are less impacted coincides with the paper by Cummins and Lewis (2003). Nevertheless, considering the mutually opposite signs of coefficients, it is hardly possible to make any consistent conclusion.

In Table 4.8 large companies yield to higher values, which might seem to be contradictory to the previous rationale. However, we also must take into account, that large companies are the reinsurance and full-line insurance ones, which will drive the statistics. Thus, the previous statements go along with our main findings.

Subsector and size split. In Table 4.8 we report the daily statistics of the Brown and Warner (1985) test for groups of companies according to their business subsectors and sizes. For discussing the business split a couple of facts need to be pointed out. Reinsurance firms generate the highest values of statistics, while P&C generate the smallest ones. This behavior is consistent with the intuition that P&C losses are typically limited by some threshold (e.g. under XL reinsurance contracts), while reinsurers step in for the larger amounts above the threshold. Furthermore, on September 17 and 18 the signs of the statistics for P&C and reinsurance are different. For the days of highest statistics of the whole set of companies the significance and signs of statistics remain the same. It is also interesting to see that for September 13 small companies experience negative and significant statistics, while middle-sized and larger companies positive ones.

Variable	September 11		September 17	
Intercept	0.27939		-0.4275	.
$\widetilde{\text{MC}}$	0.61292	*	-0.6476	*
SUB (Baseline: P&C)				
Full-line	-0.06829		0.2949	
Reinsurance	-2.04886	***	1.8401	***
$\widetilde{\text{GPWL}}$	-0.21973		0.4154	*
$\widetilde{\text{GPWNL}}$	-0.87502	**	0.8561	**
	$R^2 = 0.7429$		$R^2 = 0.7339$	

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.7: Estimated coefficients, significance and R^2 of the regression models for the dates of September 11 and September 17.

In Table 4.8 large companies yield higher values, which might seem to be contradictory to the previous rationale. However, we also must consider that large companies are the reinsurance and full-line insurance ones, which will drive the statistics. Thus, the previous statements go along with our main findings.

Market models. We now apply two other (less advanced and sophisticated) models namely market-adjusted returns and mean-adjusted returns models (cf. Equations 15 and 16) and compare them to the SIMM. The summary is presented in Table 4.9. The statistics and significance are similar for SIMM and mean-adjusted returns models, while the market-adjusted returns model generates less significant statistics. The market-adjusted returns model assumes the expected returns to be equal to market index returns. Using a broad market index (here: the global STOXX Global 1800) this approach weakly predicts companies' returns. Thus, differences between this model and SIMM can be neglected. In our further analyses the SIMM is used as standard because it is more stable and like the mean-adjusted one.

Length of the estimation window. So far, we have set the length of the estimation window to $\Delta = 120$, based on the choice by Angbazo and Narayanan (1996). Other event studies in the insurance context use longer time periods. For example, Lamb (1998) and Lamb (1995) use 150 trading days, Shelor et al. (1992) use 200. Thus, to verify that the length of the estimation period does not affect our conclusions, we report the output of the Brown and Warner (1985) test for $\Delta \in \{50, 100, 150, 200\}$ in Table 4.10. Indeed, the statistics and the obtained significance levels are the same for all values of Δ.

Market index. Now we investigate the sensitivity of the results to changes in the market index. In a *ceteris paribus* analysis we keep all parameters fixed and only vary the market index. Table 4.11 details the Brown and Warner (1985) test statistics and significances for different indices. The correspondent full names of indices' tickers can be found below the table. From Table 4.11 it can be observed, that using specific insurance indices (SXIE, BINE and SINLIE) produces more significance and higher value of statistics than broader indices. In fact, this observation goes in contrast with intuition: the insurance companies' stock prices should deviate

		Overall			FL			P&C			Re			Small			Middle			Large		
Date	W.day	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif
09-11	Tues	-5.624	-11.229	***	-4.915	-8.425	***	-2.394	-2.375	**	-16.715	-20.209	***	-1.905	-2.857	***	-7.691	-9.678	***	-11.254	-13.169	***
09-12	Wed	-3.664	-7.314	***	-1.852	-3.175	***	-8.641	-8.573	***	1.080	1.306		-5.549	-8.321	***	-1.815	-2.284	**	-1.511	-1.768	*
09-13	Thurs	-0.286	-0.571		1.073	1.839	*	-3.256	-3.231	***	1.367	1.653		-2.719	-4.077	***	-0.474	-0.596		4.745	5.552	***
09-14	Fri	-3.097	-6.184	***	-2.367	-4.058	***	-2.951	-2.928	***	-6.567	-7.939	***	-1.999	-2.997	***	-4.359	-5.484	***	-4.191	-4.904	***
09-17	Mon	0.673	1.344		0.828	1.419		-2.284	-2.266	**	7.409	8.958	***	-2.335	-3.502	***	2.726	3.429	***	4.894	5.726	***
09-18	Tues	-0.512	-1.022		-0.537	-0.921		0.312	0.310		-2.465	-2.980	***	0.550	0.825		-4.254	-5.353	***	0.639	0.748	
09-19	Wed	-1.061	-2.118	**	-1.086	-1.862	*	-1.418	-1.406		-0.062	-0.075		-1.458	-2.186	**	1.261	1.587		-2.299	-2.691	***
09-20	Thurs	-5.064	-10.111	***	-4.675	-8.013	***	-4.265	-4.232	***	-8.717	-10.539	***	-4.175	-6.261	***	-4.876	-6.136	***	-7.006	-8.198	***
09-21	Fri	-4.292	-8.568	***	-4.854	-8.321	***	-2.690	-2.669	***	-5.906	-7.141	***	-3.055	-4.582	***	-7.660	-9.638	***	-3.817	-4.467	***
09-24	Mon	3.496	6.981	***	4.409	7.558	***	0.794	0.788		6.374	7.706	***	1.707	2.559	**	5.192	6.532	***	5.593	6.544	***
09-25	Tues	1.573	3.141	***	1.324	2.270	**	1.820	1.806	*	2.014	2.435	**	1.552	2.327	**	0.355	0.447		2.683	3.140	***
09-26	Wed	2.475	4.942	***	1.697	2.908	***	2.178	2.161	**	6.528	7.893	***	1.459	2.188	**	3.747	4.715	***	3.396	3.973	***
09-27	Thurs	0.400	0.799		0.448	0.768		-0.373	-0.370		2.132	2.577	**	0.189	0.283		-0.432	-0.543		1.552	1.816	*
09-28	Fri	1.437	2.869	***	0.586	1.004		1.995	1.979	*	3.658	4.423	***	2.362	3.541	***	0.553	0.695		0.361	0.423	

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.8: Brown and Warner (1985) test statistics for different groups of companies.

		SIMM			Market-adjusted			Mean-adjusted		
Date	W.day	$\bar{A}_t$, %	Stat.	Signif.	$\bar{A}_t$, %	Stat.	Signif.	$\bar{A}_t$, %	Stat.	Signif.
09-11	Tues	-5.624	-11.229	***	-3.639	-3.610	***	-6.502	-10.267	***
09-12	Wed	-3.664	-7.314	***	-3.963	-3.931	***	-3.539	-5.589	***
09-13	Thurs	-0.286	-0.571		-0.355	-0.353		-0.262	-0.414	
09-14	Fri	-3.097	-6.184	***	-1.948	-1.932	*	-3.608	-5.698	***
09-17	Mon	0.673	1.344		2.907	2.884	***	-0.314	-0.495	
09-18	Tues	-0.512	-1.022		-0.046	-0.045		-0.723	-1.142	
09-19	Wed	-1.061	-2.118	**	-0.275	-0.273		-1.413	-2.230	**
09-20	Thurs	-5.064	-10.111	***	-2.896	-2.873	***	-6.022	-9.509	***
09-21	Fri	-4.292	-8.568	***	-3.562	-3.533	***	-4.619	-7.293	***
09-24	Mon	3.496	6.981	***	1.119	1.110		4.532	7.156	***
09-25	Tues	1.573	3.141	***	1.185	1.175		1.737	2.742	***
09-26	Wed	2.475	4.942	***	2.512	2.492	**	2.452	3.872	***
09-27	Thurs	0.400	0.799		-0.587	-0.582		0.826	1.305	
09-28	Fri	1.437	2.869	***	-0.721	-0.715		2.376	3.752	***

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.9: Brown and Warner (1985) test statistics and significance for the SIMM, the market-adjusted and the mean-adjusted return models.

stronger from broader indices, rather than insurance specific ones. Nevertheless, it seems that insurance-specific indices reject null hypothesis very frequently.

4.5 Application to other events

In this section, we present and discuss event study results for hurricanes, earthquakes, winter storms and airline crashes separately. Following the findings from the previous section, we will focus on using the Brown and Warner (1985) test statistic.[10]

4.5.1 Hurricanes event study

Hurricane Katrina. We start our set of analyses with the 2005 Atlantic hurricane season, which resulted in 3 913 fatalities and 159.2 billion (2005 USD) total damage. This season includes three hurricanes from Swiss Re sigma report "The 40 most costly insurance losses (1970-2014)". Since hurricanes Katrina, Rita and Wilma occurred with intervals of no more than one month, we use the same estimation window for each of them, namely, from March 7 to August 22. The beginning of Hurricane Katrina's event window is set to August 23, which is the start of the hurricane's lifetime. Katrina made its first landfall in U.S. state of Florida on August 25, and later in Louisiana on August 29. The end of the life cycle was on August 31. The end of two calendar weeks is on September 5, which is national holiday in US (Labor Day). Thus, we limit the event window to September 2.[11]

For the sample of all North American firms, the statistic is significant only on August 30 for

[10] Results from other tests and cross-sectional regression analyses are reported in the Appendix.

[11] In their study, Gangopadhyay et al. (2010) use a larger event window, ending on 13 September. Due to the fact that we have found no significance for the extended event window and, in order to use the same framework to all events, we stick to our initial two weeks event window.

		$\Delta = 50$			$\Delta = 100$			$\Delta = 150$			$\Delta = 200$		
Date	W.day	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif
09-11	Tues	-5.750	-10.742	***	-5.724	-11.984	***	-5.692	-9.317	***	-5.907	-9.932	***
09-12	Wed	-3.446	-6.438	***	-3.611	-7.560	***	-3.611	-5.911	***	-3.617	-6.081	***
09-13	Thurs	-0.103	-0.192		-0.248	-0.520		-0.245	-0.402		-0.272	-0.457	
09-14	Fri	-3.098	-5.787	***	-3.141	-6.576	***	-3.121	-5.109	***	-3.259	-5.481	***
09-17	Mon	0.510	0.952		0.557	1.166		0.593	0.970		0.354	0.596	
09-18	Tues	-0.410	-0.765		-0.510	-1.068		-0.500	-0.818		-0.576	-0.968	
09-19	Wed	-1.007	-1.880	*	-1.080	-2.262	**	-1.065	-1.744	*	-1.171	-1.968	*
09-20	Thurs	-5.218	-9.747	***	-5.176	-10.836	***	-5.141	-8.416	***	-5.373	-9.035	***
09-21	Fri	-4.229	-7.900	***	-4.307	-9.018	***	-4.293	-7.028	***	-4.393	-7.387	***
09-24	Mon	4.026	7.521	***	3.688	7.721	***	3.657	5.987	***	3.843	6.462	***
09-25	Tues	1.804	3.370	***	1.632	3.417	***	1.630	2.669	***	1.633	2.746	***
09-26	Wed	2.642	4.936	***	2.506	5.247	***	2.510	4.109	***	2.474	4.159	***
09-27	Thurs	0.721	1.347		0.499	1.045		0.489	0.800		0.546	0.919	
09-28	Fri	1.934	3.612	***	1.614	3.378	***	1.586	2.597	**	1.752	2.945	***

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

Table 4.10: Brown and Warner (1985) test statistics for different estimation window lengths Δ.

the Brown and Warner 1985 test (which is also justified by modified rank test). The value of the statistic is negative; therefore, we conclude that Hurricane Katrina negatively influences the given set of North American companies. Furthermore, in the frame of North American companies, the sample of small firms, as well as Full Line insurers do not show any response to this event. Small companies are less exposed to the loss from this hurricane and investors believe to diversification of risks of Full Line insurance companies. Reinsurance companies show the lowest abnormal return -1.25% on August 30, which makes this group the most affected.

For the whole set of Western European companies, the Brown and Warner 1985 test statistics is significant only on August 23. We consider this behavior as an outlier. Small-sized, large-sized and Full-Line insurers are not affected, while reinsurance companies, as well as middle-sized, reacted positively only on September 1. The small companies probably have not been exposed to the damage of this hurricane, while at the same time large companies are supposed to be stable to such events. In contrast, European P&C insurers show negative significance on 23 August. It is very unlikely that European insurers foresee the extent of loss for this hurricane. We report no effect for overall group of European insurers and a weak positive effect on reinsurance companies.

Hurricane Rita. The next major hurricane during 2005 Atlantic hurricane season is Hurricane Rita, which had a lifetime between September 19 and September 26. Again, to be consistent we extend the event widow to keep ten trading days and set the ending date to September 30. The chosen event window is smaller than in Gangopadhyay et al. (2010), but we apply the same reasoning as above. Hurricane Rita made its landfall in Louisiana on September 24 (Saturday), which might be incorporated in the nearby days.

In contrast to Hurricane Katrina, Rita generated more dates with significant abnormal returns, which are also higher in absolute values. There are no subgroups of North American companies with less than two significant dates, and significant dates correspond to each other for different groups. For the overall North American sample, the first two significant dates prior the landfall are negative, while the one right before the landfall is positive. Finally, the last significant date is again positive. In other words, for early days of formation the effect is negative, but

		BKXE Index			SXXP Index			SXFINE Index			SXIE Index			BINE Index			SINLIE Index		
Date	W.day	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif	$\bar{A}_t$, %	Stat	Signif
09-11	Tues	-4.262	-8.928	***	-4.272	-8.798	***	-3.119	-6.609	***	-2.657	-5.703	***	-2.509	-5.486	***	-2.453	-5.335	***
09-12	Wed	-3.826	-8.015	***	-4.218	-8.685	***	-3.816	-8.087	***	-3.372	-7.237	***	-3.355	-7.334	***	-3.377	-7.346	***
09-13	Thurs	-0.517	-1.084		-0.595	-1.225		-0.864	-1.830	*	-1.322	-2.838	***	-1.362	-2.978	***	-1.427	-3.103	***
09-14	Fri	-1.650	-3.458	***	-1.822	-3.751	***	-1.315	-2.786	***	-1.705	-3.660	***	-1.624	-3.551	***	-1.622	-3.528	***
09-17	Mon	-1.072	-2.246	**	-1.108	-2.282	**	-1.005	-2.129	**	-1.298	-2.786	***	-1.319	-2.884	***	-1.383	-3.009	***
09-18	Tues	-0.523	-1.095		-0.421	-0.867		-0.265	-0.562		-0.640	-1.374		-0.620	-1.356		-0.766	-1.667	*
09-19	Wed	-0.675	-1.414		-0.575	-1.183		-0.408	-0.864		-0.874	-1.876	*	-0.867	-1.896	*	-0.936	-2.036	**
09-20	Thurs	-4.481	-9.387	***	-4.663	-9.602	***	-3.724	-7.892	***	-3.817	-8.193	***	-3.730	-8.156	***	-3.789	-8.240	***
09-21	Fri	-3.657	-7.661	***	-3.701	-7.620	***	-3.201	-6.783	***	-3.097	-6.647	***	-3.008	-6.576	***	-3.042	-6.617	***
09-24	Mon	2.479	5.194	***	2.728	5.617	***	1.811	3.837	***	2.141	4.595	***	2.047	4.476	***	2.127	4.627	***
09-25	Tues	1.271	2.662	***	1.285	2.646	***	0.572	1.212		0.810	1.739	*	0.771	1.686	*	0.785	1.708	*
09-26	Wed	1.848	3.872	***	1.987	4.093	***	1.781	3.773	***	1.641	3.522	***	1.609	3.517	***	1.587	3.451	***
09-27	Thurs	0.430	0.900		0.319	0.656		-0.003	-0.007		-0.066	-0.141		-0.087	-0.191		-0.098	-0.213	
09-28	Fri	1.299	2.722	***	1.194	2.459	**	1.161	2.461	**	1.158	2.485	**	1.120	2.449	**	1.168	2.540	**

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests.

BKXE Index = EURO STOXX Total Market
SXXP Index = STOXX Europe 600
SXFINE Index = EURO STOXX Financials
SXIE Index = EURO STOXX Insurance
BINE Index = EURO STOXX Total Market Insurance
SINLIE Index = EURO STOXX Total Market Nonlife Insurance

Table 4.11: Brown and Warner (1985) test statistics for different market indices.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
08-23	Tues	0.519	0.502	0.661	0.180	0.936	-0.994	0.015	-1.679*	-1.419	-0.558	-1.265	-1.806*	-1.103	-0.402
08-24	Wed	-0.162	0.278	-0.315	-0.462	-0.042	-0.702	0.084	-0.195	-0.267	-0.341	0.356	-0.375	-0.092	0.238
08-25	Thurs	1.368	1.212	1.833*	0.531	1.647	-0.001	0.616	-0.882	0.669	-1.366	-1.470	0.130	-1.348	-0.525
08-26	Fri	-0.725	-1.245	0.242	-0.510	-0.680	-0.863	-0.154	0.455	0.247	0.508	0.104	0.387	0.092	0.945
08-29	Mon	-0.655	0.116	-0.439	-1.351	-0.449	-0.198	-1.587	-1.605	-0.989	-1.361	-0.837	-1.343	-1.111	-1.192
08-30	Tues	-2.120**	-1.571	-2.151**	-1.684*	-2.081**	-0.749	-2.208**	-0.371	-0.419	0.638	-1.075	-0.533	-0.147	-0.052
08-31	Wed	0.097	1.541	-0.913	-0.815	0.577	-1.017	-0.824	0.261	0.154	0.496	-0.216	0.224	0.806	-1.634
09-01	Thurs	0.631	0.508	0.550	0.522	0.765	0.468	-0.313	1.243	-0.045	2.254**	0.294	-0.224	1.443	2.159**
09-02	Fri	-0.788	-1.149	-1.310	0.389	-0.971	0.457	-0.810	-0.645	-0.354	-0.716	-0.148	-0.257	-0.368	-1.342

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.12: BW1985 statistics and significance for Hurricane Katrina.

days around and after landfall are positive. It can be rationally explained by the presence of the fear of further damages and/or even stronger consequences as of Hurricane Katrina for the early part. The later positive response is likely due to more precise and optimistic prediction of weaker damages, than from Hurricane Katrina.[12] The split according to the size does not show critical differences, while the subsector split shows that for reinsurance firms the effect is stronger.

Western European companies show almost the same pattern: a negative response before the landfall, which is later replaced by a positive one. However, for the overall group the positive effect was delayed to the consequent trading day – September 26 (Monday). Small- and middle-sized companies do not generate any significance, while large companies have five significant dates with the pattern described above. The group of P&C companies also has no significant dates. Furthermore, the most affected group are reinsurance companies.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
09-19	Mon	-2.188**	-2.658***	-0.978	-1.499	-1.816*	-1.429	-2.993***	-0.705	-0.175	0.713	-2.362**	0.336	-1.161	-0.724
09-20	Tues	-0.306	-0.359	-0.167	-0.200	-0.225	-0.542	-0.131	0.633	0.261	0.195	0.914	0.329	0.735	0.057
09-21	Wed	-3.615***	-1.950*	-4.023***	-3.422***	-3.294***	-0.554	-5.808***	-2.265**	-1.248	-0.763	-2.711***	-1.779*	-1.703*	-1.469
09-22	Thurs	-0.845	-1.414	-0.617	0.034	-0.688	-0.692	-1.046	-1.580	-0.033	-0.859	-2.755***	-0.397	-1.181	-3.025***
09-23	Fri	3.435***	2.469**	3.892***	2.510**	2.697***	2.092**	5.595***	1.296	0.783	-0.212	2.260**	1.495	0.396	1.387
09-26	Mon	-0.184	-0.204	1.271	-1.158	-0.589	-0.091	1.718*	1.947*	0.043	1.562	2.759***	0.654	1.383	3.552***
09-27	Tues	0.230	-0.337	0.034	0.902	-0.175	1.737*	0.058	-0.443	-0.301	-0.328	-0.219	-0.610	-0.426	0.598
09-28	Wed	-0.825	-1.413	-1.495	0.737	-0.814	0.128	-1.347	1.084	0.404	0.863	0.966	0.555	0.978	0.925
09-29	Thurs	1.746*	1.350	1.656	1.409	2.046**	0.477	0.462	-1.392	-0.593	-1.392	-0.770	-0.037	-2.083**	-0.393
09-30	Fri	-0.198	-0.724	-0.015	0.353	-0.134	-0.297	-0.205	1.176	1.178	0.114	0.953	1.443	0.597	0.369

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.13: BW1985 statistics and significance for Hurricane Rita.

Hurricane Wilma. Hurricane Wilma formed on October 16 and dissipated on October 27, making its landfall in Mexico on October 21 and in Florida on October 24. Our results indicate only positive significant statistics for the overall group of North American insurance companies. The stocks of large-sized companies generate the largest values and number of significant statistics (as well as the cross-sectional mean of abnormal returns) among the groups according to the size. For this group the only significant and negative statistic appears on October 20, one day before landfall in Mexico. In contrast, middle-sized companies show the least number of significance, with only one significant and positive statistics on October 24. For small-sized

[12] `http://www.washingtonpost.com/wp-dyn/content/article/2005/09/24/AR2005092401758.html`

companies, statistics are positive and significant on October 19 and 24, which seems to be a mix of large- and middle-sized companies' daily statistics. As for previous events, Full Line insurance firms are not significantly affected, while reinsurance companies show the highest values of means and statistics. It is interesting to see, that for Western European companies the effect is mixed: on October 19, the statistic is negative and significant, and on October 24 it is positive. The most affected group of companies is the middle-sized ones. P&C and Full Line insurance firms show similar results, which coincides with one for overall group.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
10-17	Mon	-1.374	-1.554	-1.489	-0.421	-1.403	-0.155	-1.465	-0.137	0.635	-0.186	-1.070	0.739	-0.672	-0.435
10-18	Tues	-1.031	-0.285	-1.101	-1.315	-0.834	-0.326	-1.940*	-0.648	-0.836	-0.658	0.506	0.234	-1.378	0.401
10-19	Wed	2.765***	2.169**	1.654	2.919***	3.146***	1.340	0.449	-3.667***	-1.433	-3.713***	-2.095**	-2.566**	-3.284***	-1.631
10-20	Thurs	-1.082	-0.846	0.113	-1.714*	-0.899	-1.496	-0.516	0.493	-0.676	1.144	0.804	-0.412	0.629	1.450
10-21	Fri	1.951*	1.353	1.634	1.911*	1.618	0.456	3.677***	-1.229	-0.713	-1.839*	0.366	-0.996	-1.179	-0.017
10-24	Mon	2.091**	2.045**	1.825*	1.326	1.776*	1.386	2.656***	2.226**	1.160	1.955*	1.251	1.705*	1.861*	1.006
10-25	Tues	-1.062	-1.202	-1.003	-0.434	-1.218	0.391	-1.234	-0.514	-1.754*	0.443	0.852	-0.991	-0.414	1.126
10-26	Wed	-0.410	-0.409	-1.018	0.242	0.024	-0.838	-1.537	0.520	0.839	0.556	-0.649	0.323	0.598	-0.005
10-27	Thurs	0.024	-0.654	0.489	0.425	-0.278	0.580	0.711	-0.852	-0.100	-1.146	-0.505	-1.325	-0.245	-0.155
10-28	Fri	1.618	0.892	0.855	2.216**	1.792*	1.102	0.103	-1.184	-0.099	-1.254	-1.111	0.247	-1.446	-1.716*

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.14: BW1985 statistics and significance for Hurricane Wilma.

Hurricane Ike. The lifetime of Hurricane Ike runs from September 1 to September 15 in 2008. It overlaps with another event relevant to insurance companies and to overall world economy, namely, the bankruptcy of Lehman Brothers Holdings Inc. On September 15, the latter filed for Chapter 11 bankruptcy protection, therefore we exclude this date from the initial event window. Moreover, we also set the beginning of the event window to September 2, because September 1 was Labor Day in the USA. The hurricane made its landfall in Cuba on September 8 and in Texas on September 13. Since September 13 fell on a Saturday, the information of the landfall in Texas is incorporated only in the Monday data of the next trading week, which is the day of Lehman Brothers bankruptcy.

Surprisingly, North American companies, as well as middle-sized and small-sized companies are not impacted by this event (no significance during the event window). The negative statistics on September 12 for large-sized companies is probably related to the bankruptcy, rather than to the hurricane (the bankruptcy likely indirectly influenced the assets of the large companies). On the same date, the negative and significant statistics appears also for Full Line insurance, as well as on September 9, after the landfall in Cuba. In contrast, reinsurance companies show the positive and significant statistics on September 8.

On September 2, Western European companies produce positive and significant statistics, and negative ones for September 10 and 11 negative ones. Companies of different groups show different results with similar behavior only on September 11 (negative and significant statistics). Thus, we can report a negative impact on European companies. However, if this event is extremely close to another major financial event, we cannot distinguish the effect of which event is dominant.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
09-02	Tues	0.600	0.515	0.760	0.323	0.604	0.528	0.451	1.823*	1.635	1.356	1.596	0.003	2.471**	2.056**
09-03	Wed	1.124	0.980	1.151	0.906	1.187	0.728	0.908	-0.855	-1.673*	-0.518	-0.112	-0.321	-0.993	-0.784
09-04	Thurs	0.043	-0.264	0.475	-0.129	-0.041	0.457	-0.109	-0.059	-0.036	0.275	-0.430	0.174	-0.087	-0.450
09-05	Fri	0.988	0.143	1.397	1.151	0.979	1.218	0.240	-1.224	-0.298	-1.332	-1.199	-0.437	-1.437	-1.203
09-08	Mon	0.170	0.313	0.162	-0.034	0.072	-0.405	1.753*	1.128	0.753	0.351	1.764*	0.727	1.111	1.007
09-09	Tues	-1.008	-0.322	-0.977	-1.499	-0.890	-1.720*	-0.167	1.272	1.291	0.870	1.071	0.709	1.312	1.192
09-10	Wed	0.492	0.903	0.348	0.048	0.607	-0.416	1.062	-1.904*	-0.227	-2.503**	-1.678*	-0.877	-2.250**	-1.228
09-11	Thurs	-1.205	-0.750	-1.032	-1.543	-1.255	-0.724	-1.186	-2.210**	-1.816*	-1.678*	-2.035**	-1.925*	-1.813*	-2.055**
09-12	Fri	-1.319	-1.414	-0.228	-2.075**	-0.583	-3.880***	-1.386	0.440	-0.055	0.660	0.385	-0.104	0.660	0.520

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.15: BW1985 statistics and significance for Hurricane Ike.

Hurricane Irene. Hurricane Irene made its landfall in the USA after Hurricane Ike in 2008. Irene originated on Sunday August 21. Thus, we adapt the event window to start from the consequent trading day, August 22. Hurricane Irene disappeared on August 30. However, due to the presence of the significant statistics of the following days, which may be related to the hurricane, we extend the event window to September 2. Hurricane Irene affected the east of the US, including the landfalls in the states North Carolina, New Jersey and New York. The landfall in this region happened on August 27 (a Saturday). The stock market exchange was not closed following the hurricane. Thus, we expect August 29 to incorporate the information about the impact of the hurricane.

The impact is similar for the North American and Western European companies. Also for both groups, the effect of the splitting in terms of the size and the business subsector is weak enough to be neglected. The reaction of companies' returns can be described in the following way: before the landfall in New Jersey and New York (August 27) the cross-sectional abnormal returns mean are negative, but no significantly different from zero. On the Monday following the landfall, when the market has already absorbed the information about losses, the statistics show positive and significant results. This may potentially be explained by the investors' expectations prior the landfall, that the hurricane would produce much higher damage. Thereafter, the abnormal returns drop to the negative values, for some groups even becoming significant. Probably, more precise information about re-estimated claims became publicly available affecting stock prices negatively.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
08-22	Mon	-0.665	-0.374	-1.040	-0.408	-0.443	-0.787	-1.411	-0.139	-1.453	0.055	0.568	-1.031	0.322	0.400
08-23	Tues	-0.509	-0.771	-0.007	-0.691	-0.556	-0.103	-0.635	-1.032	0.433	-1.511	-1.151	-0.400	-1.739*	0.385
08-24	Wed	0.404	0.036	0.651	0.422	0.371	0.793	0.041	0.794	0.212	0.569	1.165	-0.575	1.486	1.122
08-25	Thurs	-1.450	-2.278**	-0.165	-1.735*	-1.534	-0.437	-1.884*	-0.289	-0.818	0.048	-0.208	-0.940	-0.112	0.703
08-26	Fri	-0.273	-0.417	-0.438	0.091	-0.247	-1.063	0.545	-1.558	-0.571	-1.045	-2.260**	-0.506	-1.593	-2.842***
08-29	Mon	3.846***	3.791***	3.131***	3.912***	3.711***	3.834***	3.660***	2.297**	3.166***	1.416	1.911*	3.127***	1.608	1.799*
08-30	Tues	-1.666*	-1.195	-1.731*	-1.737*	-1.502	-1.940*	-1.790*	-0.881	-1.402	-0.186	-0.999	-0.469	-1.089	-0.632
08-31	Wed	-1.576	-1.975*	-1.487	-0.988	-1.778*	-0.947	-0.993	0.453	-0.789	0.632	0.937	-0.572	1.086	0.419
09-01	Thurs	-2.286**	-2.800***	-2.066**	-1.593	-2.349**	-2.793***	-0.948	-0.233	1.015	-0.151	-1.074	-0.343	-0.199	0.018
09-02	Fri	-0.577	-0.623	-0.328	-0.687	-0.387	-0.892	-1.001	-0.871	-0.294	-0.536	-1.331	-0.526	-0.833	-1.247

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.16: BW1985 statistics and significance for Hurricane Irene.

Hurricane Sandy. We set the event window to start on October 22 and to end on November 2. This period exactly coincides with the hurricane's lifetime. Hurricane Sandy, the second-costliest hurricane in the United States history, made its main damage in New Jersey and New York from October 28 to October 30. The New York Stock Exchange and Nasdaq were closed during October 29 and 30, and reopened on October 31, so that we exclude these dates from the event window.

The group of North American companies reacted with significantly negative statistics on the last day of the event window – November 2. While small-sized companies are not impacted, the middle- and large-sized firms show similar results as for the overall group with extra significant and negative statistics for large-sized companies on October 26. There are no critical differences between the responses of P&C and Full Line insurance companies. In contrast, reinsurance companies yield five consequent significant dates: positive on October 25 and negative for the remaining dates in the event window.

For most groups of Western European companies the effect is absent, except small-sized companies and reinsurance companies. For the case of small-sized companies this is most probably linked to the random occurrence (small European companies are very unlikely exposed to the influence of the US hurricane), while, given the extent of this hurricane, the strong negative impact on reinsurers goes along with basic intuition of high losses of these companies.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
10-22	Mon	0.817	0.616	0.633	0.645	0.761	0.490	0.756	0.878	1.829*	0.151	0.744	0.567	0.875	0.778
10-23	Tues	-0.482	-0.553	-0.106	-0.100	-0.394	-0.895	-0.093	-0.727	-0.406	-0.591	-0.711	-0.321	-0.826	-0.805
10-24	Wed	0.357	0.024	0.861	0.468	0.195	0.852	0.711	-0.160	-0.035	-0.071	-0.298	-0.339	0.146	-0.659
10-25	Thurs	0.051	0.226	-0.358	-0.156	-0.220	0.251	2.271**	0.202	0.062	0.233	0.122	0.351	-0.034	0.392
10-26	Fri	-0.766	0.049	-1.479	-1.756*	-0.419	-1.343	-2.219**	-0.520	-0.406	-0.325	-0.591	-0.539	-0.258	-0.760
10-31	Wed	-0.492	-0.238	-0.536	-0.669	-0.380	0.085	-1.748*	-0.132	-0.735	-0.059	0.316	-0.199	-0.270	1.168
11-01	Thurs	-1.007	-0.699	-0.574	-1.232	-0.801	-0.495	-2.400**	-0.958	-2.199**	0.217	-1.064	-1.213	-0.359	-0.908
11-02	Fri	-2.589**	-1.230	-2.333**	-4.148***	-1.955*	-3.438***	-4.011***	-0.191	0.390	-0.113	-0.652	-0.202	-0.020	-0.713

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.17: BW1985 statistics and significance for Hurricane Sandy.

Comments. We do not find a unique reaction pattern of stock prices. Among the different hurricanes, the behavior of price is relatively heterogeneous. In most cases, the evolution of the stock market reaction depends on the coverage and representation by the media. Furthermore, we find the responses of North American and Western European companies to differ significantly. In several instances, this might be linked to the distance to the catastrophe. Often, the most affected group are reinsurance companies.

4.5.2 Earthquakes event study

Earthquake in Chile. The 2010 Chile earthquake had a magnitude of 8.8, which makes it the seventh largest earthquake since 1900 in terms of magnitude and the eleventh in terms of property damages. The earthquake also triggered a tsunami, which affected the state of California and the Japanese region Tōhoku. On March 10, Swiss Re published a press release[13] giving an estimation of the loss. This date is included into the event window.

[13] http://www.swissre.com/media/news_releases/provisional_estimates_of_losses_from_Chilean_EQ_and_European_storm.html

We observe no significant impact of the earthquake in Chile on the overall group of North American companies, as well as for small-sized, large-sized and the P&C group of firms. Middle-sized insurance companies show a negative result on the first trading day after the event (March 1), and a positive result on March 10 (the date of press release). It might be explained by the overestimated losses at the beginning, and followed by the positive reaction after the reestimation. For Full Line insurance, we report only positive influence, while the impact on reinsurance companies is negative (two negative and significant statistics on March 5 and 9).

The companies with origin in Western Europe generate positive and significant statistics on March 10. Western European companies show different results depending on their size: small companies show positive (significant) statistics on March 10, middle-sized ones show in addition positive (significant) statistics on March 2. Finally, for large companies, statistics are firstly negative (March 1), then positive (March 3), and almost at the end of the estimation window (March 9) again negative. The P&C, Full Line insurers and reinsurers produce positive and significant statistics on March 2, 3 and 10, respectively.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
03-01	Mon	-1.205	-0.430	-1.935*	-0.927	-1.275	-0.654	-0.900	-1.506	-0.107	-1.529	-1.951*	-0.947	-1.380	-1.492
03-02	Tues	0.964	1.308	0.869	0.101	1.158	0.943	-1.034	1.379	1.377	1.741*	0.179	2.315**	0.561	0.348
03-03	Wed	0.624	0.923	0.300	0.210	0.666	0.093	0.933	1.371	0.565	0.814	1.987**	1.249	0.883	1.764*
03-04	Thurs	0.220	-0.105	-0.073	0.814	-0.005	0.946	-0.010	-0.593	-0.402	-0.945	-0.035	-0.658	-0.237	-0.999
03-05	Fri	-0.168	0.553	-0.826	-0.421	0.040	0.052	-1.825*	-0.286	0.504	-0.807	-0.284	-0.104	-0.334	-0.206
03-08	Mon	0.466	0.152	0.388	0.717	0.221	0.943	0.678	0.587	0.418	0.600	0.397	-0.172	0.686	1.379
03-09	Tues	-0.729	-0.845	-0.467	-0.445	-0.800	0.484	-2.175**	-0.399	0.636	0.156	-1.787*	-0.892	0.357	-1.447
03-10	Wed	1.099	0.490	1.794*	0.690	0.947	1.741*	-0.151	1.903*	1.859*	1.738*	1.042	1.086	2.120**	0.787
03-11	Thurs	0.437	0.421	0.366	0.299	0.603	-0.048	0.013	-0.171	0.099	-0.076	-0.438	-0.274	0.001	-0.332
03-12	Fri	-0.247	-0.324	0.234	-0.469	-0.300	-0.223	0.246	0.881	0.254	0.578	1.317	0.100	1.135	0.798

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.18: BW1985 statistics and significance for the earthquake in Chile.

Christchurch earthquake. The earthquake in Christchurch occurred on February 22 and had a magnitude of 6.3. Nevertheless, the earthquake aftermath was two times more expensive than the earthquake in Chile in terms of insured losses. We limit the event window by March 4. Such groups of North American companies, as the overall, small-sized, large-sized, P&C and Full Line insurers do not seem to be affected, while reinsurers and middle-sized are affected negatively. The negative and significant statistics for both groups can be observed on February 22, when the earthquake happened. Our results for European companies show a negative effect for reinsurance on Wednesday, March 2. The small-sized companies show positive with a subsequent negative reaction. However, this behavior is probably not related to the earthquake. These companies are likely to have only very limited exposure in New Zealand.

Tōhoku earthquake. The 2011 Tōhoku earthquake (Great East Japan earthquake) is the fourth most powerful earthquake since 1900, and the strongest in Japan. Furthermore, the earthquake produced losses to make it the second costliest catastrophes in the time range between 1970 and 2016. The earthquake triggered a tsunami, which at the same time caused a nuclear accident in Fukushima Daiichi nuclear power plant. The earthquake occurred on March 11 with a magnitude of 9.0. Takao et al. (2013) studied the impact of this earthquake on five

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
02-22	Tues	-1.200	-0.240	-1.777*	-1.409	-0.914	-0.961	-2.592**	-0.274	-0.688	-0.108	-0.014	-0.878	0.122	0.225
02-23	Wed	-0.333	-0.123	-0.571	-0.235	-0.251	-0.152	-0.891	-0.636	-0.580	-0.547	-0.500	-0.680	-0.432	-0.628
02-24	Thurs	0.251	1.185	-0.332	-0.339	0.331	0.123	-0.145	-0.792	-0.292	-0.540	-1.082	-0.566	-0.970	0.059
02-25	Fri	0.167	-0.156	0.681	-0.046	0.117	-0.522	1.316	0.907	0.256	1.201	0.725	0.156	1.362	0.374
02-28	Mon	0.172	-0.519	0.657	0.448	0.034	-0.158	1.333	1.060	2.415**	0.305	0.348	1.132	0.897	0.378
03-01	Tues	-1.465	-1.275	-1.412	-1.336	-1.535	-0.741	-1.366	0.303	0.922	0.619	-0.597	0.048	0.381	0.417
03-02	Wed	-0.229	0.178	-0.619	-0.240	-0.315	0.036	-0.000	-1.282	-1.683*	-0.565	-1.152	-0.876	-1.016	-2.112**
03-03	Thurs	1.376	1.094	1.453	1.243	1.331	1.352	1.020	0.085	0.228	-0.423	0.441	0.568	-0.078	-0.759
03-04	Fri	-0.761	-0.801	-0.745	-0.509	-0.781	-0.031	-1.286	0.926	0.350	1.279	0.623	0.624	0.796	1.318

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.19: BW1985 statistics and significance for Christchurch earthquake.

Japanese companies, choosing the length of the event window equal to 15 days, ending on April 4. We stick to our initial length of ten trading days, which includes the important dates of two press releases by Munich Re on March 14 and March 22, and the press release of Swiss Re on March 21, each one related to claims estimation. We set the estimation window to be the same as for the earlier earthquake in Christchurch to avoid overlapping.

For all groups of companies from North America, except reinsurance companies, the event produced no effect according to the BW1985 test. Further, the returns of reinsurance companies show a negative response on the day of the event and a positive one on March 21 (press release of Swiss Re). In contrast, Western European companies are more sensitive to Tōhoku earthquake. The overall group shows two negative and significant statistics on March 11 and 15, and then two positive ones on March 17 and 21. The split according to the size only shifts the significant dates, but does not change the pattern "negative-then-positive". P&C companies do not generate any significant dates, while Full Line insurance companies do on the same dates as for overall group. Finally, for the reinsurers group the significant and negative dates are March 11 and 14 and positive on March 17.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
03-11	Fri	0.210	0.532	-0.388	0.407	0.452	1.074	-2.453**	-2.561**	-0.649	-3.098***	-2.391**	-1.485	-1.887*	-5.504***
03-14	Mon	0.848	0.602	0.814	0.946	0.710	1.270	0.658	-1.229	-1.338	-0.727	-1.127	-0.646	-0.934	-2.722***
03-15	Tues	0.790	0.934	0.813	0.358	0.759	1.141	0.113	-1.716*	-3.242***	-1.092	-0.474	-1.149	-2.175**	0.184
03-16	Wed	-0.797	-0.678	-0.801	-0.710	-0.501	-1.198	-1.560	-1.036	0.074	-1.027	-1.440	-0.748	-1.274	0.119
03-17	Thurs	0.104	0.093	-0.065	0.274	0.250	-0.136	-0.449	2.025**	-0.350	2.431**	2.555**	1.153	1.685*	3.687***
03-18	Fri	0.635	0.006	1.150	0.667	0.801	-0.816	1.383	0.180	-0.451	1.148	-0.374	0.048	0.422	-0.545
03-21	Mon	1.244	0.860	1.175	1.435	0.975	1.623	1.670*	1.829*	2.353**	0.523	1.958*	0.964	2.270**	0.724
03-22	Tues	-0.710	-0.315	-1.006	-0.671	-0.666	-0.462	-0.965	-0.557	-1.106	-0.154	-0.310	-1.140	0.066	-0.682
03-23	Wed	-1.261	-1.488	-1.186	-0.696	-1.315	-1.184	-0.463	0.417	2.691***	-1.088	0.021	0.707	0.400	-0.770
03-24	Thurs	-0.341	-0.642	-0.114	-0.118	-0.281	-0.119	-0.828	0.506	0.158	0.325	0.732	-0.011	0.912	-0.086

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.20: BW1985 statistics and significance for Tōhoku earthquake.

Comments. For the considered earthquakes, we observe the absence of stock market significant reactions of North American firms. Western European companies' stock prices respond more significantly to Christchurch and Tōhoku earthquakes. Thus, we assume that European companies (and the reinsurers within that group) are more exposed to the earthquake risk from the investors' perspective. In fact, as in the case of the hurricanes, the reinsurance subsector is found to be the most affected.

4.5.3 Storms event study

Winter storm Kyrill. Winter storm Kyrill caused damages mostly in Great Britain and Germany. Despite the fact it was formed on January 15, the storm reached Great Britain only on January 17. Most losses in Germany have occurred during the period from January 17 to January 19. The storm dissipated on January 24, but we set the end of the event window on January 23. Munich Re published its press release[14] on January 19, which compared this winter storm to Daria and Lothar winter storms.

The overall group of North American companies reacted negatively on January 22, which was mostly due to the contribution from middle-sized companies, since neither small-sized, nor large-sized companies have significant dates. Furthermore, P&C companies do not show significant dates, while Full Line insurer and reinsurers produce negative responses.

The group of Western European companies did not respond to Kyrill, as well as the group of small-sized and large-sized. The middle-sized companies produced a negative significant statistic on January 23. Full Line insurance companies do not show any significant statistics. Finally, P&C and reinsurers demonstrate negative results on March 19 and 22, respectively.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
01-17	Wed	-0.954	-0.582	-0.986	-0.970	-0.923	-0.799	-0.730	-0.470	-0.602	-0.038	-0.409	-0.577	-0.393	0.124
01-18	Thurs	-0.473	-0.092	-0.897	-0.426	-0.430	-0.532	-0.398	0.046	0.616	-0.224	-0.299	0.777	-0.245	-0.891
01-19	Fri	-0.395	0.160	-0.874	-0.612	-0.070	-0.692	-1.662*	-0.408	0.208	-1.167	0.091	-2.058**	0.758	0.695
01-22	Mon	-1.748*	-0.933	-2.357**	-1.499	-1.181	-1.762*	-3.775***	-1.051	0.076	-1.311	-1.091	-0.643	-0.593	-1.968*
01-23	Tues	1.183	1.537	0.030	0.986	1.076	0.752	1.442	-1.597	-0.693	-2.280**	-0.508	-1.962*	-0.576	-1.622

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.21: BW1985 statistics and significance for winter storm Kyrill.

Winter storm Klaus. The lifetime of the winter storm Klaus was from January 23 to January 28 in 2009, affecting mostly Spain and France. The storm reached Italy, but without implying large damages. Klaus caused two times fewer insured losses than the winter storm Kyrill.

There are no groups among the North American insurers that have any significance. Thus, we conclude no effect on North American companies. As expected, the overall group of Western European companies shows significance: on January 23 a negative one, and on January 26 and 28 a positive one. The middle-sized companies show the same pattern. It is interesting to see, that small-sized companies generate only negative significance, while the group of large-sized firms only positive. The groups according to the subsector of the companies have different dates with significant statistics, even though these dates correspond to the dates for overall group.

4.5.4 Airline crashes event study

Malaysia Airlines Flight 17. The Malaysia Airline Flight 17 was downed on July 17, 2014 under unknown circumstances. It is the deadliest airline shoot down incident. According to

[14] https://www.munichre.com/en/media-relations/publications/press-releases/2007/2007-01-19-press-release/index.html

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
01-23	Fri	-0.358	-0.860	-0.114	0.014	-0.567	0.607	-0.611	-2.099**	-1.914*	-1.729*	-1.484	-2.319**	-1.203	-2.394**
01-26	Mon	0.083	0.699	-0.198	-0.294	0.128	-0.048	-0.004	2.420**	1.389	2.046**	2.221**	2.670***	1.839*	1.413
01-27	Tues	0.119	0.096	-0.168	0.388	0.165	0.399	-0.739	-0.472	0.326	0.116	-1.268	-0.136	-0.423	-0.694
01-28	Wed	0.640	0.014	0.995	0.768	0.539	0.612	0.750	2.223**	0.579	1.788*	2.488**	1.416	2.046**	1.917*

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.22: BW1985 statistics and significance for winter storm Klaus.

The Wall Street Journal[15] most insurers have stated they would still cover the private claims even though that the crash might be treated as an act of the war in the Russian-Ukrainian conflict.

For the overall group of North American companies, we observe no effect, as well as for small- and middle-sized firms. For large-sized companies the statistic is significant and negative on July 22, which is also supported by the rank test. Thus, we cannot ignore the negative effect on large firms. Among groups by business split, only Full Line insurers are negatively impacted on July 22. The negative significant statistic of BW1985 test for reinsurance firms is not supported by rank test, thus we ignore it and refer to some random occurrence. Western European firms react negatively on July 18. However, only groups of middle-sized and Full Line insurance firms have significant and negative statistics on that day.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
07-17	Thurs	-1.027	-0.861	-0.517	-1.319	-0.827	-0.645	-1.840*	0.067	0.061	0.457	-0.514	0.559	-0.217	-0.217
07-18	Fri	0.823	0.806	0.846	0.323	0.760	0.949	0.452	-1.962*	-1.263	-2.152**	-1.319	-1.552	-1.885*	-1.034
07-21	Mon	0.056	-0.094	0.024	0.326	0.016	-0.072	0.383	-0.552	-0.787	0.053	-0.626	-1.316	0.137	-0.530
07-22	Tues	-1.464	-0.411	-0.819	-3.219***	-1.438	-1.720*	-0.331	-0.178	-0.022	-0.446	0.086	-0.066	-0.543	1.118
07-23	Wed	0.232	-0.214	0.570	0.465	0.147	0.117	0.648	0.226	0.110	0.050	0.479	0.287	-0.058	0.879

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.23: BW1985 statistics and significance for Malaysia Airlines Flight 17.

Germanwings Flight 9525. On March 24, 2015, because of the suicide of the co-pilot, the aircraft of Germanwings flight 9525 crashed in the French Alps, killing all 150 people aboard. Allianz SE, the leading insurer of the all-risk policy, estimated the losses from the Germanwings accident amounts to $300 million[16]. AIG and Lufthansa's captive insurer Delvag are also involved to the claim payments[17] together with a group of more than 30 other insurers.

Expectedly none of the North American insurance companies shows any significant abnormal returns. At the same time, Western European companies generate a negative and significant statistic on March 26, two days after the crash. The significance of the BW1985 test is supported by the rank test. Both middle- and large-sized companies contribute to the significance of this date to the overall group, while the small-sized companies' stock prices stay indifferent. We observe that the stocks of Western European P&C companies do not react to the event,

[15] http://www.wsj.com/articles/more-insurers-to-waive-right-to-dismiss-claims-in-flight-17-crash-1407319040
[16] http://www.insuranceinsider.com/germanwings-loss-reserve-set-at-300mn
[17] http://www.insuranceinsider.com/allianz-leads-germanwings-loss

while Full Line insurers and reinsurers do.

Date	W.day	North America							Western Europe						
		Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re
03-24	Tues	-0.366	0.063	-0.340	-0.922	-0.256	-0.516	-0.707	1.260	1.515	0.478	1.227	0.964	1.166	1.089
03-25	Wed	-0.274	-0.466	-0.277	0.235	-0.500	-0.488	1.269	1.455	1.621	0.727	1.398	1.452	1.224	0.924
03-26	Thurs	-0.662	-0.251	-1.468	-0.005	-0.876	0.259	-0.282	-1.862*	-0.928	-2.047**	-1.690*	-1.149	-1.780*	-2.084**
03-27	Fri	-0.515	-0.248	-0.608	-0.552	-0.583	-0.325	-0.197	-0.562	-0.274	-0.687	-0.429	-0.998	-0.227	-0.322
03-30	Mon	1.427	1.246	1.509	0.795	1.585	0.832	0.809	0.069	-0.083	-0.460	0.889	-0.834	0.399	0.882

*, **, *** stand for statistical significance at the 10%, 5%, 1% percent levels for two-sided tests;
Ov. = Overall, S = Small, M = Middle, L = Large;
P&C = Property & Casualty, FL = Full Line, Re = Reinsurance.

Table 4.24: BW1985 statistics and significance for Germanwings Flight 9525.

4.6 Summary and concluding remarks

This paper examines the impact of 13 natural and man-made catastrophes on 87 listed insurance companies. The summary is presented in Table 4.25. We do not find any clear pattern of the stock price behavior in any type of the catastrophes. Despite the fact, that we find no significance of geographical origin coefficient in regression, we observe that North American companies are more influenced by the local events, such as hurricanes on US territory and experience almost no effect from external ones. In contrast, European companies respond to all events, including international ones. This may be linked to the finding that reinsurance companies are more sensitive to the occurrence of catastrophes compared to the P&C firms. Typically, the catastrophe risk of P&C companies is covered by reinsurance treaties (e.g., non-proportional excess of loss) that limit their losses. At the same time, reinsurance companies experience losses when such catastrophes appear. Finally, based on the obtained results, we cannot state that investors discriminate and distinguish companies according to any of the studied characteristics.

Event name	North America							Western Europe							Significant characteristic
	Ov.	S	M	L	P&C	FL	Re	Ov.	S	M	L	P&C	FL	Re	
Hurricane Katrina	-	0	+/-	-	-	0	-	0	0	+	0	0	0	+	SUB(Re)
Hurricane Rita	-/+	-/+	-/+	-/+	-/+	+	-/+	-/+	0	0	-/+	-	-	-/+	$\widetilde{MC}$, $\widetilde{GPWL}$
Hurricane Wilma	+	+	+	+/-	+	0	+	-/+	-	-/+	-	-/+	-/+	-	GEO
Hurricane Ike	0	0	0	-	0	-	+	+/-	-	-	+/-	-	+/-	+/-	$\widetilde{GPWL}$
Hurricane Irene	+/-	-/+	+/-	-/+	+/-	+/-	-/+	+	+	0	-/+	+	-	-/+	
Hurricane Sandy	-	-	-	-	-	+/-	0	+/-	0	0	0	0	-		
Earthquake in Chile	0	0	-/+	0	0	+	-	+	+	+	-/+	+	+	+	SUB(Re), $\widetilde{GPWL}$
Christchurch earthquake	0	0	-	0	0	0	-	0	+/-	0	0	0	0	-	SUB(FL), $\widetilde{GPWL}$
Tōhoku earthquake	0	0	0	0	0	0	-/+	-/+	-/+	-/+	-/+	0	-/+	-/+	SUB(FL)
Winter storm Kyrill	-	0	-	0	0	-	-	0	0	-	0	-	0	-	
Winter storm Klaus	0	0	0	0	0	0	0	-/+	-	-/+	+	-/+	+	-/+	
Malaysia Airlines Flight	0	0	0	-	0	-	-	-	0	-	0	0	-	0	GEO/WE
Germanwings Flight	0	0	0	0	0	0	0	-	0	-	-	0	-	-	$\widetilde{MC}$, $\widetilde{GPWNL}$

0 = no effect
\- = negative effect
\+ = positive effect
+/- = mixed effect, with firstly positive significance
-/+ = mixed effect, with firstly negative significance

Table 4.25: Summary of the impact of the selected events on insurance companies.

Methodological comments. We apply a variety of tests, both parametric and nonparametric, to examine the performance of the event study methodology. The test proposed by Brown and Warner (1985) show the highest robustness and stability among others. In addition, we show that the approach is relatively insensitive to changes in the length of the estimation window. We also vary the market model and the market index for the SIMM to select the optimal ones. Since several approaches lead to similar results, the reliability of the event study method seems high in the considered reference event. However, we find contradicting results in the regression analyses, which leaves a doubt on their appropriateness of application to our setup. In fact, the number of companies in the different cells (size, business split, geography) is relatively small diminishing the reliability. The heterogeneity and the various parameters that can be varied in the event study method may also explain the lack of clear response pattern found along the different catastrophe events.[18] These remarks outline the importance and relevance of expert opinion for prospective applications in the insurance industry.

Findings for the insurance industry. Our observations add to the understanding how the insurance market stock valuation behaves and what type of events may be significantly influential. Applying event study analyses may help firms to diversify their risk exposure and to improve their investment and crisis management strategies. For instance, the difference in responses of North American and Western European companies provides evidence for the diversification across the geographical origins in the portfolio selection. However, the absence of strong significance in many statistics support the assumption that from the view of investors the risk of catastrophes is incorporated into the prices of contracts. In other words, a certain number of (enormous) losses caused by catastrophes are expected and reflected in the insurance companies' strategies.

Open points. Although the presented event study analyses are comprehensive, our research results leave several questions unanswered. A main direction of further research could be to include several other parameters, such as a geographical split of revenues to the regression analysis of CAR. This would allow a more precise definition of the exposure to certain catastrophes, and derive whether investors make their decisions based on company-specific information. Research could also extend on testing the reaction to regulatory events or other insurance-linked events, which do not lead to direct losses. Finally, since media coverage has proven to be an indicator for stock responses, one could focus on the relationship between stock behavior and news releases.

Bibliography

Angbazo, L. A. and R. Narayanan, 1996, Catastrophic Shocks in the Property-Liability Insurance Industry: Evidence on Regulatory and Contagion Effects, *Journal of Risk and Insurance*, 63(4):619–637.

Boehmer, E., J. Musumeci, and A. B. Poulsen, 1991, Event-Study Methodology Under Conditions of Event-Induced Variance, *Journal of Financial Economics*, 30(2):253–272.

Brown, S. J. and J. B. Warner, 1980, Measuring Security Price Performance, *Journal of Financial Economics*, 8(3):205–258.

Brown, S. J. and J. B. Warner, 1985, Using Daily Stock Returns: The Case of Event Studies, *Journal of Financial Economics*, 14(1):3–31.

[18] See the electronic Appendix available from the author for all test results and regression analyses.

Campbell, J., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*.

Chen, X., H. Doerpinghaus, B.-X. Lin, and T. Yu, 2008, Catastrophic Losses and Insurer Profitability: Evidence From 9/11, *Journal of Risk and Insurance*, 75(1):39–62.

Chesney, M., G. Reshetar, and M. Karaman, 2011, The impact of terrorism on financial markets: An empirical study, *Journal of Banking and Finance*, 35(2):253–267.

Corrado, C. J., 1989, A Nonparametric Test for Abnormal Security-Price Performance in Event Studies, *Journal of Financial Economics*, 23(2):385–395.

Corrado, C. J. and T. L. Zivney, 1992, The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns, *Journal of Financial and Quantitative Analysis*, 27(3):465–478.

Cowan, A. R., 1992, Nonparametric Event Study Tests, *Review of Quantitative Finance and Accounting*, 2(4):343–358.

Cummins, J. and C. Lewis, 2003, Catastrophic Events, Parameter Uncertainty and the Breakdown of Implicit Long-Term Contracting: The Case of Terrorism Insurance, *Journal of Risk and Uncertainty*, 26(2/3):153–178.

Doherty, N. A., J. Lamm-Tennant, and L. T. Starks, 2003, Insuring September 11th: Market Recovery and Transparency, *Journal of Risk and Uncertainty*, 26(2-3):179–199.

Ewing, B. T., S. E. Hein, and J. B. Kruse, 2006, Insurer Stock Price Responses to Hurricane Floyd: An Event Study Analysis Using Storm Characteristics, *Weather and Forecasting*, 21(3):395–407.

Fama, E. and K. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33:3–56.

Gangopadhyay, P., J. D. Haley, and L. Zhang, 2010, An Examination of Share Price Behavior Surrounding the 2005 Hurricanes Katrina and Rita, *Journal of Insurance Issues*, 33(2):132–151.

Grace, M. F., J. Rauch, and S. Wende, 2014, Systemic Risk and Interconnectedness in the Financial Industry: Implications on Regulation of Financial Conglomerates.

Henderson, G. V. J., 1990, Problems and Solutions in Conducting Event Studies, *Journal of Risk and Insurance*, 57(2):282–306.

Kolari, J. W. and S. Pynnonen, 2011, Nonparametric rank tests for event studies, *Journal of Empirical Finance*, 18(5):953–971.

Kothari, S. P. and J. B. Warner, 2007, Econometrics of Event Studies, In B. E. Eckbo, editor, *Handbook of Corporate Finance, Volume 1*, chapter 1, pages 4–36. Elsevier.

Lamb, R. P., 1995, An Exposure-Based Analysis of Property-Liability Insurer Stock Values Around Hurricane Andrew, *Journal of Risk and Insurance*, 62(1):111–123.

Lamb, R. P., 1998, An Examination of Market Efficiency Around Hurricanes, *Financial Review*, 33(1):163–172.

Marlett, D. C., R. B. Corbett, and C. Pacini, 2000, Insurer Stock Price Responses to the Disclosure of Revised Insured Loss Estimates After the 1994 Northridge Earthquake, *Journal of Insurance Issues*, 23(2):103–123.

Marlett, D. C., J. Griffith, C. Pacini, and R. E. Hoyt, 2003, Terrorism Insurance Coverage: The Market Impact on Insurers and Other Exposed Industries, *Journal of Insurance Regulation*, 22(2):41–62.

Patell, J. M., 1976, Corporate Forecasts of Earnings Per Share and Stock Price Behaviour: Empirical Tests, *Journal of Accounting Research*, 14(2):246–276.

Shelor, R. M., D. C. Anderson, and M. L. Cross, 1990, The Impact of the California Earthquake on Real Estate Firms' Stock Value, *Journal of Real Estate Research*, 5(3):335–340.

Shelor, R. M., D. C. Anderson, and M. L. Cross, 1992, Gaining from Loss: Property-Liability Insurer Stock Values in the Aftermath of the 1989 California Earthquake, *Journal of Risk and Insurance*, 59(3):476–488.

Shelor, R. M. and M. L. Cross, 1990, Insurance Firm Market Response to California Proposition 103 and the Effects of Firm Size, *Journal of Risk and Insurance*, 57(4):682–690.

Takao, A., T. Yoshizawa, S. Hsu, and T. Yamasaki, 2013, The Effect of the Great East Japan Earthquake on the Stock Prices of Non-life Insurance Companies, *The Geneva Papers on Risk and Insurance Issues and Practice*, 38(3):449–468.

Thomann, C., 2013, The Impact of Catastrophes on Insurer Stock Volatility, *Journal of Risk and Insurance*, 80(1):65–94.

Wang, Y. and R. B. Corbett, 2008, Market Efficiency: Evidence from Market Reactions of Insurance Industry Stocks to the September 11, 2001 Event, *Journal of Insurance Issues*, 31(2):152–167.

Wilcoxon, F., 1945, Individual Comparisons by Ranking Methods, *Biometrics Bulletin*, 1(6):80–83.

Yamori, N. and T. Kobayashi, 2002, Do Japanese Insurers Benefit from a Catastrophic Event?, *Journal of the Japanese and International Economies*, 16(1):92–108.

Yanase, N. and Y. Yasuda, 2010, The Impact of the September 11 Terrorist Attack on the Global Insurance Markets: Evidence from the Japanese Property-Casualty Insurance Industry, *Journal of Insurance Issues*, 33(1):85–107.

Chapter 5

A Note on Stochastic Programming for the Asset Allocation in Swiss Pension Funds

The aim of this research is to study the properties of stochastic programming for the asset allocation in a framework of Swiss pension funds under solvency constraints. We develop a simplified yet scalable model using a linear utility function as an objective function. Further, we use a vector autoregressive model as the underlying econometric model for the assets and calibrate it using monthly returns of bonds and stock indices. In our case, the stochastic programming problem cannot be solved exactly, that is why the deterministic equivalent represented by a scenario tree is used. The scenario tree is built by discretizing the econometric model using the "bracket-mean" method. In this paper, we empirically study the convergence of the initial asset allocation with respect to the number of scenarios. We observe the results for the probability of deficit, as well as for the expected value of the deficit given shortage. Further, we test the sensitivity of the above-mentioned characteristics to changes in internal parameters of the model, namely, planning horizon, target wealth and shortage penalty. Finally, the effect of misestimating the stock volatility and the bonds expected return on the initial asset allocation is examined.

Note: This is joint work with J. Wagner. The authors are thankful for the comments on earlier versions of this manuscript from participants in the following conferences: R in Insurance (Paris, June 2017) and 21st International Congress on Insurance: Mathematics and Economics (Vienna, July 2017).

5.1 Introduction

The occupational pensions in Switzerland (second pillar) have been mandatory since 1985. Together with the state pension (first pillar) it aims to attain a salary replacement ratio of 50 to 70%. The number of such pension funds has dramatically decreased due to the complexity of the regulation and thus has been outsourced to larger groups (see, e.g., Bütler and Ruesch, 2007, for an overview of the pension system in Switzerland). At the genesis of the second pillar, a majority of pension funds used a defined benefit (DB) scheme. However, over time the number of pension funds using a defined contribution (DC) plan dramatically replaced DB plans. According to Bütler and Ruesch (2007) in 2015 the proportion of DC pension funds reached 85%: "from 2005 to 2015 the number of pension funds with defined benefit plans fell steadily from 289 to 58" (Swiss Federal Statistical Office, 2017). However, DC plans in Switzerland significantly differ from those in the rest of the world, and should rather be considered as a hybrid of DC and DB and typically referred as to cash-balanced plan.[1]

In 2015 the total asset of the occupational pension was 123% of Swiss GDP (CHF 787 bn or USD 793 bn), which makes the asset allocation of pension funds of high importance. Two strategic oppositely directed goals of pension funds are long-term safety, to cover deferred liabilities, and generating high returns. To assess the solvency level and the performance of a pension fund, the funding ratio is a key measure: It is defined as the proportion of the total assets over the total liabilities. This ratio should be above 100% for a healthy pension fund. In 2015, the average funding ratio of 29 Swiss Market Index companies' pension funds was 83% (Willis Tower Watson, 2016), which is considered rather low. The real net rate of investment returns in 2015 was 2.2%. The proportion invested in equities increased from 24.6% (2005) to 29.8% in 2015, while shares invested in bonds decreased from 37.2% to 32.9% (OECD, 2015). Investing more aggressively, that is with a larger share invested in highly volatile securities typically increases the average return, but also increase the associated risk.

Traditionally, immunization methods, namely cash-flow and duration matching, are used for determining the asset allocation. However, these approaches focus only on fixed-income securities, which in a low interest rate environment generate too low returns. To address this issue the modern portfolio theory by Markowitz (1952) allows incorporating stochastic returns, and therefore, using equities as alternative asset class. This model and its extensions (e.g., liabilities-driven investments, surplus optimization) have the disadvantage of being limited to a single-period framework. That this issue is typically solved by using dynamic investment strategies, such as buy & hold, constant mix and CPPI (see Perold and Sharpe, 1988, for an overview). However, these methods do not yield an optimal allocation, such as stochastic control does. By utilizing stochastic differential equations (SDE), the optimal asset allocation over a time span can be obtained. Difficulties in SDE coefficients' estimations, challenges with implementing complex regulatory restrictions and other limitations make this method less attractive for practical use. Stochastic programming is considered as an alternative to stochastic control methods. It exploits the so-called scenario tree concept to deal with randomness, while optimizing the objective function. Furthermore, it supports the use of various constraints. One of the first successful applications of stochastic programming as an ALM tool is the Russell-Yasuda Kasai model for a Japanese P&C insurer (Cariño and Ziemba, 1998; Cariño et al., 1994; and Cariño

[1] In an international perspective all Swiss pension plans, DB and DC, considering the guaranties provided, are often classified as defined benefits plans.

et al., 1998). Applications are not limited only to the insurance industry: among many others, Kusy and Ziemba (1986) developed a model for the Vancouver City Savings Credit Union, Koivu et al. (2005) present a model for a Finnish pension fund, Geyer and Ziemba (2008) introduce InnoALM, an Austrian pension fund financial planning model. Last but not the least, Monte-Carlo (MC) simulation methods are extensively used in practice. The latter two methods are considered in this paper.

We present a simplified example of a stochastic programming (SP) model for a pension fund. The SP problems should be discussed in conjunction with the concept of scenario trees. The scenario tree is based on an underlying economic model, which is the distribution of asset returns. We use a vector-autoregressive model (VAR) arguing our choice as the trade-off between complexity and precision. We generate the scenario tree by a “mean-bracket” method. Using monthly historical returns data, we calibrate the VAR model by estimating coefficients. Based on the reference model we study the convergence of the optimal asset allocation at an initial time with respect to the number of scenarios, which defines how detailed the approximation of the stochastic process will be. Further, we analyze the sensitivity of the optimal allocation, as well as several key performance indicators (e.g., the expected deficit and the probability of the deficit) to changes in certain model parameters, namely: planning horizon, target wealth (or total value of liabilities), shortage penalty (the parameter of the utility function), securities' expected returns and volatilities. The given model is also compared with the result of Monte-Carlo simulation methods. Our main findings are as follows: (1) the SP approach is very attractive from a theoretical perspective, but very demanding in terms of computational power and mathematical involvement, (2) the optimal solution converges to its exact value with respect to the scenario tree discretization parameter.

The remainder is organized as follows. In Section 5.2, we present the reference problem formulation and key performance characteristics. Then, we describe the VAR as underlying economic model. In the end of the chapter, we outline a “bracket-mean” method for scenario tree generation. In Section 5.3, we present numerical illustrations. We calibrate the model on the monthly returns and study the properties of the reference case, namely the convergence with respect to “bushiness” (the number of successors for each intermediate node) of the scenario tree and the sensitivity to changes of the model parameters. In Section 5.4, we draw our concluding remarks.

5.2 Model framework

In this section, we describe an extensible and scalable simple stochastic programming problem, that describes some of a pension fund's strategic objectives. The pension fund can reallocate its assets. By letting the shares invested in different asset classes be decision variables, we derive an optimal asset allocation for meeting the pension fund's goals linked to an objective function. The objective function is chosen to be a utility function described by the amount of deficit and amount of surplus. Stochastic assets' returns are modeled by VAR and the scenario tree generation method is chosen to be “bracket-mean”.

5.2.1 Stochastic programming for pension funds

Stochastic programming has a long history of application as a method for asset allocation and ALM in pension funds. The existing literature can be conceptually divided into two streams of research. The first stream considers different ways of modeling the pension funds, including a vast number of theses and papers. In his thesis, Dert (1995) considers a model for DB scheme pension fund using chance constraints. Drijver (2005) focuses on particularity of pension funds in the Netherlands. Other similar works include Reynisson (2012), Dondi (2005) and Toukourou (2008). Hilli et al. (2007) develops a model for a Finnish pension fund. The second stream of papers studies the mathematical aspects of SP problems, for example, the scenario generation methods. The contribution of Bogentoft et al. (2001) is to use CVaR as a risk measure (their SP problem minimizes the costs of a fund while fixing the conditional value at risk). Fleten et al. (2002) examine SP with respect to fixed mix portfolio models. Kouwenberg (2001) compares several scenario tree generation techniques while minimizing the sum of the average contribution rates staying in the framework of a Dutch pension fund. Høyland and Wallace (2001) also consider the issue of scenario tree generation. Kaut and Wallace (2007) use a more systematic approach for evaluating the quality of various scenario tree generating methods. Further examples of authors contributing to the scenario tree techniques are: Høyland and Wallace (2001), Pennanen and Koivu (2005), Dupacová et al. (2000), Hochreiter and Pflug (2007), Heitsch and Römisch (2009).

SP, as with any other mathematical optimization problems, consists of two main parts, an objective function and the constraints. These two parts determine the strategical goal of the pension fund, as well as the variety of restrictions and requirements imposed by the regulatory authorities. Common ways of optimizing objective functions are maximizing the total value of assets, maximizing the expected value of the utility function, maximizing the funding ratio, minimizing the contribution rate or the capital injection (in case of defined-benefits pension schemes).

One can argue that maximizing the total value of assets without any risk constraints implies risk neutrality from the investor. This can be solved by using a utility function as an objective function. As alternative one can maximize the total value of assets controlling the probability of deficit (chance constraints) or the average shortage given deficit (integrated chance constraints). It is possible to express both chance constraints and integrated chance constraints in linear form. However, the former one will require variables to be binary (which is the particular case of integer restriction) that involves mixed integer programming (see, e.g., Toukourou and Dufresne, 2018).

In order to make the problem more flexible and easier to extend we remain in a linear framework. It entails the linearity of the objective function as well as constraints. Thus, we are limited to use the simplest linear utility functions. Kallberg and Ziemba (1983) discuss various utility functions applied to the financial engineering.

Uncertainty in the model is represented by introducing the multistage scenario tree concept. Our model maximizes the value of the utility function only at the end of the planning horizon. The structure of the scenario tree is defined in the constraints matrix by budget constraints. Budget constraints ensure that full wealth is invested in asset classes for each node and provide non-anticipative decisions. It is also possible to apply other constraints, for instance, portfolio constraints ensuring that the share invested in a certain asset class is in line with regulation (see, e.g., Hilli et al., 2007), transaction constraints restricting upper bounds of purchases or sales. We omit listed constraints, as well as transaction costs in order to simplify the model.

5.2.2 Reference case

Our model is based on standard procedures presented for example in Birge and Louveaux (2011, pp. 20-27) and Shapiro et al. (2009, p. 14), and adapted then to a pension fund. We consider a time span of T years ($t = 0, 1, ..., T$). At the beginning of the year 0 the pension fund has an amount A_0 of assets. The pension fund's goal is to meet its liability L_T (target wealth) at the end of the planning horizon T. For simplicity, we ignore (neglect) the intermediate years' aims at $t = 1, ..., T-1$. In our model, we consider two asset classes available for investment, namely stocks and bonds, with returns $r_{1,t}$ and $r_{2,t}$, respectively. These returns are random, which leads to a stochastic programming problem formulation. The pension fund is allowed to reallocate its assets at each beginning of a year, i.e. at times $t = 0, ..., T-1$. The goal of the pension fund is expressed as a simple linear utility function maximized at time T:

$$U(A_T) = q \cdot (A_T - L_T)^+ - r \cdot (L_T - A_T)^+, \tag{56}$$

where q is an excess (surplus) reward, r is a deficit (shortage) penalty, and $(x)^+ = \max(x, 0)$ is the positive part function. This function can be easily interpreted. Missing the target wealth forces the pension fund to borrow at costs r the deficit, while outperforming would lead to an additional profit q on the surplus.

We let L_T be a pre-defined parameter of the utility function (i.e., no fluctuations or risk until time T) and the asset part A_T depend on both random returns and decision variables linked to the asset allocation. At this point, we introduce notations for scenarios and decision variables (see Section 5.2.4 for the scenario tree definition and the generation procedure). Let s_t indicate a scenario at time $t \in \{1, 2, ..., T\}$, and $S_t = \{s_t^1, s_t^2, ..., s_t^m\}$ represents a set of possible values of s_t, given a particular history of path $s_1 = s'_1, s_{t-2} = s'_{t-2}, ..., s_{t-1} = s'_{t-1}$. Further, let $x_{1,t}(s_1 = s'_1, ..., s_t = s'_t)$ and $x_{2,t}(s_1 = s'_1, ..., s_t = s'_t)$ for $t = 0, ..., T-1$ be decision variables, which indicate the amount of money units invested in bonds and stocks, respectively, given a history $s_1 = s'_1, ..., s_t = s'_t$. For simplicity, we use short-hand notations, namely $\boldsymbol{s}'_t = (s_1 = s'_1, ..., s_t = s'_t)$. Then, the decision variables are $x_{1,t}(\boldsymbol{s}'_t)$ and $x_{2,t}(\boldsymbol{s}'_t)$, where the prime sign indicates a particular path of scenarios from the sets $S_1, ..., S_t$. The same system of notations is applied to returns, that is $r_{1,t}(s_1 = s'_1, ..., s_t = s'_t)$ and $r_{2,t}(s_1 = s'_1, ..., s_t = s'_t)$ for $t = 1, ..., T$ are bonds' and stocks' returns, respectively, given a history $s_1 = s'_1, ..., s_t = s'_t$ (shorter notation equivalents are $r_{1,t}(\boldsymbol{s}'_t)$ and $r_{2,t}(\boldsymbol{s}'_t)$).

Using such notation, we can present budget constraints for the initial assets:

$$A_0 = x_{1,0} + x_{2,0}, \tag{57}$$

and for intermediate time points $t \in 0, ..., T-2$ and for scenarios $s_{t+1} \in S_{t+1}$ schematically represented on Figure 5.1, we have:

$$x_{1,t}(\boldsymbol{s}'_t) \cdot e^{r_{1,t+1}(\boldsymbol{s}'_t, s_{t+1})} + x_{2,t}(\boldsymbol{s}'_t) \cdot e^{r_{2,t+1}(\boldsymbol{s}'_t, s_{t+1})} = x_{1,t+1}(\boldsymbol{s}'_t, s_{t+1}) + x_{2,t+1}(\boldsymbol{s}'_t, s_{t+1}). \tag{58}$$

In order to calculate the value of the utility function at time T for each scenario, values of deficit and excess must be computed. For this purpose, we use additional decision variables, the amount of deficit $w(s_1, ..., s_T)$ and the amount of surplus $y(s_1, ..., s_T)$. Then, we can define additional constraints at T for scenarios $s_T \in S_T$ to compute $w(s_1, ..., s_T)$ and $y(s_1, ..., s_T)$ schematically depicted in Figure 5.2:

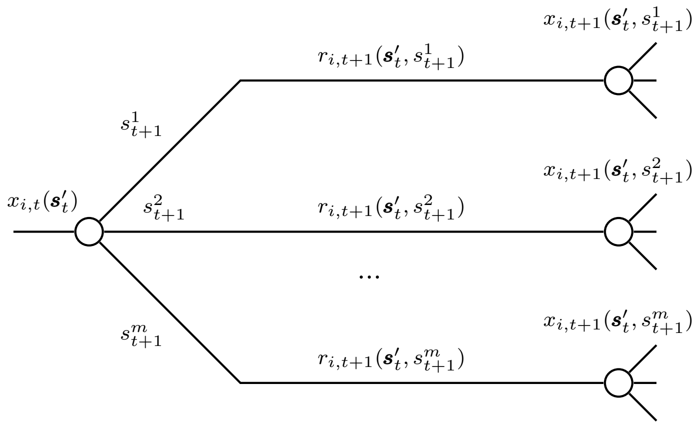

Figure 5.1: Illustration of an intermediate step at time t in a scenario tree.

$$L_T = x_{1,T-1}(\boldsymbol{s}'_{T-1}) \cdot e^{r_{1,T}(\boldsymbol{s}'_{T-1}, s_T)} + x_{2,T-1}(\boldsymbol{s}'_{T-1}) \cdot e^{r_{2,T}(\boldsymbol{s}'_{T-1}, s_T))} - y(\boldsymbol{s}'_{T-1}, s_T) + w(\boldsymbol{s}'_{T-1}, s_T). \quad (59)$$

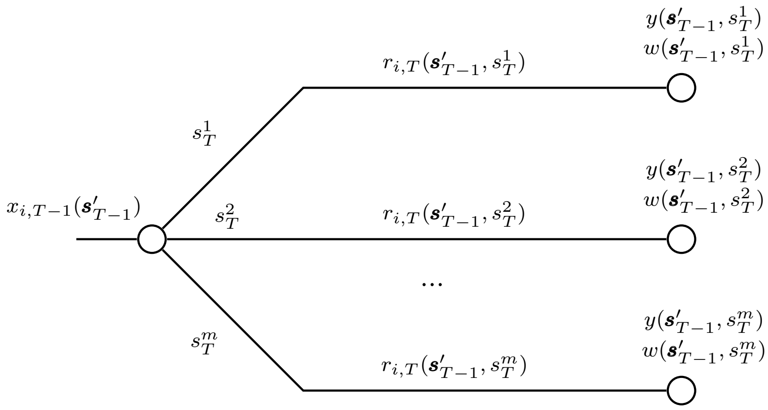

Figure 5.2: Illustration of the final step at time $T-1$ in a scenario tree.

Using above formulas, the expected utility function can be expressed as:

$$\mathbb{E}(U(A_T)) = \sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} (q \cdot y(s'_1, ..., s'_T) - r \cdot w(s'_1, ..., s'_T)) \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T). \quad (60)$$

In our further analysis, we investigate the properties of the model, namely the sensitivity of the optimal solution to changes of initial parameters, as well as the convergence of the solution with respect to the number of scenarios. Even the trivial case where we consider the dependence between the optimal solutions and the parameters of the problem is not obvious. The traditional sensitivity analysis of a linear program would be a naïve approach, because in this case we can only examine the impact of particular elements of a matrix A of constraints coefficients, a vector of right-hand constraints b and cost vector c (see Bertsimas and Tsitsiklis, 1997, for the definitions) on the optimal solution. However, these matrices are unique for each particular set of parameters. In other words, for instance, the increase of the stocks' variance will change the matrix A completely (and not just one element) with all its corresponding values. Traditional sensitivity analysis would focus only on single nodes' return values. To study these dependencies we use *ceteris paribus* sensitivity analyses by varying a particular parameter while fixing the rest. The number of asset allocations of interest growths exponentially with the planning horizon T, changing the dimensions of the matrix A. Furthermore, it is not clear how to compare two sets of asset allocations (optimal solutions) of different number of scenarios with the same parent (and thus the same history). Thus, we focus on the following values, where the last four are aggregations of the final outcomes:

- The optimal investment (share) in bonds at the initial time $t = 0$:

$$\frac{x_{1,0}}{x_{1,0} + x_{2,0}}. \tag{61}$$

- The probability of the deficit:

$$\mathbb{P}(A_T - L_T < 0) \approx \sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{w(s'_1,...,s'_T)>0\}} \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T). \tag{62}$$

- The expected value of the deficit (given shortage):

$$\mathbb{E}(A_T - L_T | A_T - L_T < 0) \approx \frac{\sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{w(s'_1,...,s'_T)>0\}} \cdot w(s'_1, ..., s'_T) \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T)}{\sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{w(s'_1,...,s'_T)>0\}} \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T)}. \tag{63}$$

- The probability of the excess:

$$\mathbb{P}(A_T - L_T > 0) \approx \sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{y(s'_1,...,s'_T)>0\}} \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T). \tag{64}$$

- The expected value of the excess (given surplus):

$$\mathbb{E}(A_T - L_T | A_T - L_T > 0) \approx \frac{\sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{y(s'_1,...,s'_T)>0\}} \cdot y(s'_1, ..., s'_T) \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T)}{\sum_{s'_1 \in S_1} \cdots \sum_{s'_T \in S_T} \mathbb{1}_{\{y(s'_1,...,s'_T)>0\}} \cdot \mathbb{P}(s_1 = s'_1, ..., s_T = s'_T)}. \tag{65}$$

We have also carried out the analysis for the above values using non-strict inequality signs and have observed that the conclusions stay the same.

5.2.3 Underlying economic models

The scenario tree must be discussed in conjunction with an underlying economic model describing the dynamics of securities' returns. Therefore, the quality of the scenario tree highly depends on how accurate and precise the economic model is. For this purpose, time series models are typically used as long as stochastic programming requires the realization of random variables only at discrete points of time.

We focus on the vector-autoregressive (VAR) model, which is a common choice in pension literature (e.g., Kouwenberg, 2001; Toukourou and Dufresne, 2018; Hilli et al., 2007). This model is the generalization of the univariate autoregressive model. The model of order p in matrix form is defined as follows:

$$\boldsymbol{r}_t = \boldsymbol{m} + \Theta_1 \boldsymbol{r}_{t-1} + \Theta_2 \boldsymbol{r}_{t-2} + ... + \Theta_p \boldsymbol{r}_{t-p} + \boldsymbol{\epsilon}_t, \tag{66}$$

where

- $\boldsymbol{r}_t = (r_{1t}, ..., r_{nt})'$ is a vector of n components (returns of the n asset classes in our case),
- $\boldsymbol{m} = (m_1, ..., m_n)'$ is a constant vector of length n, which represents the means of corresponding asset classes,
- $\Theta_1, ..., \Theta_p$ are $n \times n$ matrices of coefficients with corresponding elements $\theta^1_{i,j}, ..., \theta^p_{i,j}$ for $i, j \in \{1, ..., n\}$, and
- $\boldsymbol{\epsilon}_t$ is an error term.

The error term must satisfy the assumption of zero expected value $\mathbb{E}(\boldsymbol{\epsilon}_t) = \mathbf{0}$ and be serially uncorrelated $\mathbb{E}(\boldsymbol{\epsilon}_t \cdot \boldsymbol{\epsilon}'_{t-k}) = 0, \forall k \neq 0$. The matrix $\mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t)$ is called covariance matrix and defines a dependence structure among error terms. Also, $\boldsymbol{\epsilon}_t$ is assumed to have a multivariate normal distribution.

The model allows only for linear dependence among n variables as well as linear serial dependence. This means that variance clustering effects cannot be incorporated in such a model. In our numerical application, we use monthly returns, which in contrast to daily data show much weaker variance clustering effects. Thus, neglecting these effects, the VAR model can be considered a reasonable choice.[2]

5.2.4 Scenario tree generation

The distinctive feature of stochastic programming when compared to deterministic one is the stochasticity of the problems' parameters. For the case of discrete random variables (with relatively small number of outcomes), the problem can be directly transformed to the deterministic equivalent, for which it is possible to find an exact optimal solution. For continuous parameters or discrete variables with a large number of outcomes, the exact solution can be found only in trivial cases. More complex problems require numerical approximations.

Since the values of assets' returns are evolving over the time and are modeled by times series, it is possible to generate a vast number of sample paths, and optimize the problem individually for each path. However, such approach leads to anticipativity, i.e. the decision is dependent not only on the information known up to the time of decision, but also on the future realizations, which are not known in practice. To cope with this issue, we exploit the conditional distribution given the information up to the decision time.

The common solution is to adopt the concept of a scenario tree. This approach assumes an iterative process. Given returns at time t, we generate a certain number of returns for time

[2]Possible extensions and more advanced and sophisticated models include, for instance, Vector Equilibrium Correction (Hilli et al., 2007), which can be reduced to VAR. If one wants to incorporate a variance clustering effect a multivariate GARCH model is a natural choice. Furthermore, to model the dependencies between variables, copulas can be used (see Kaut and Wallace, 2011). The GARCH and copula models could be combined together which, however, would lead to high complexity.

$t+1$ (child nodes) with positive probabilities.

For this set of child nodes the optimal asset mix is required to be the same (as long as they have a unique predecessor), by which we satisfy the nonanticipativity constraint (for more detailed overview see Shapiro et al., 2009 and Birge and Louveaux, 2011). We repeat this procedure for each of the child nodes further in time, i.e. we generate a certain number of possible outcomes, until the required time T is attained. Unfortunately, scenario trees have an undesirable feature called "curse of dimensionality", i.e. the size of the tree exponentially grows with the number of nodes per child, which is the number of scenarios. Other methods exist, such as, for instance, hybrid models proposed by Hibiki (2006) (and references therein) and extensions proposed by Bogentoft et al. (2001).

There are four common methods of generating the scenario tree: *sampling* method (crude sampling, using variance reduction techniques and quasi-random sampling), *discretizing* method ("bracket-median" and "bracket mean"), *moment matching* method (via integration quadratures and minimizing a distance between statistical properties of the original distribution and its discrete approximation counterpart) and *"optimal discretization"* (i.e. constructing scenario trees from an auxiliary non-linear optimization problem that minimizes the distance in objective functions). In this paper we focus on the "bracket-mean" method. The idea of this discretization method is to represent the continuous (or discrete with vast number of outcomes) distribution by its discrete approximation with a limited number of possible outcomes (see Keefer and Bodily, 1983 for the univariate case). That approximation must be as close as possible to the original distribution. In the framework of this method, we split the sample space into a number of intervals (hypercubes in case of a random vector) that are called brackets. Then, the values from each bracket are represented by a pair of outcome and probability to occur of such outcome. The outcome is computed as a conditional expectation (a vector of conditional expected values in case of a random vector).

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, ..., \epsilon_{n,t})'$ be a continuous random vector with support $\mathbb{R}^n$. Suppose that we want to have k intervals per variable, then in total the set $\mathbb{R}^n$ should be partitioned into k^n parts. In order to do so, we use $k+1$ marginal quantiles $(0, \frac{1}{k}, \frac{2}{k}, ..., \frac{k-1}{k}, 1)$ of each variable in the random vector. These sets form the boundaries of the k^n hypercubes. Each of these k^n hypercubes is represented by a pair of probability and outcome, which defines the discrete probability law. The probabilities of these intervals are calculated by integrating the multivariate density function over each of the hypercube j:

$$pr_j = \int_{l_{n,j}}^{u_{n,j}} \cdots \int_{l_{1,j}}^{u_{1,j}} f(x_1, \ldots, x_n) dx_1 \ldots dx_n,^3 \quad j = 1, \ldots, k^n \tag{67}$$

where

- pr_j is the probability of occurrence of the outcome in the j-th hypercube,
- $l_{1,j}, \ldots, l_{n,j}$ and $u_{1,j}, \ldots, u_{n,j}$ are the corresponding boundaries of the j-th hypercube, which are represented by marginal quantiles, and
- $f(x_1, \ldots, x_n)$ is the joint multivariate probability density function of the vector $\boldsymbol{\epsilon}_t$.

The outcomes are represented by the mean vectors, for which the i-th component of the outcome is:

$$bm_{i,j} = \frac{1}{pr_j} \int_{l_{n,j}}^{u_{n,j}} \cdots \int_{l_{1,j}}^{u_{1,j}} x_i \cdot f(x_1, \ldots, x_n) dx_1 \ldots dx_n,^{3} \qquad (68)$$
$$j = 1, \ldots, k^n, \quad i = 1, \ldots, n.$$

For more details, Miller III and Rice (1983) and Smith (1993) cover the univariate case of the "bracket-mean" method. Miller III and Rice (1983) also report systematic underestimation of all even moments in the "bracket-mean" method, as well as on the typical underestimation or overestimation of odd moments depending on the domain of a random variable.

5.3 Numerical illustrations

In this section, we introduce historical data of monthly returns and the descriptive statistics. This data is used to calibrate the VAR model. We also highlight a particular example of discretization. Further, we present our convergence and sensitivity analyses and we compare SP and MC methods.

5.3.1 Parameters of the reference case

For the reference case, we fix the parameters as shown in Table 5.1. The choice of the utility function parameters is based on Birge and Louveaux (2011, pp. 20-27).

Parameter	Value
Initial wealth	$A_0 = 1\,000$ (money units)
Target wealth	$L_T = 1\,000$ (money units)
Surplus reward	$q = 1$
Shortage penalty	$r = 4$

Table 5.1: Parameters of the reference case.

Historical returns and calibration. In our study, we consider two exemplary asset classes namely bonds and stocks. Bonds are represented by Swiss Bond Index (Domestic Government Total Return Index) while stocks are expressed by Swiss Performance Index. For both we obtain time series of monthly returns in the period between January 1996 and December 2015 yielding to 240 observations. Table 5.2 presents the descriptive statistics.

	Bonds	Stocks
Mean return	0.0034	0.0071
Std. deviation	0.0108	0.0444
Correlation matrix		
Bonds	1	-0.1803
Stocks	-0.1803	1

Table 5.2: Mean, standard deviation and correlation structure for bonds and stocks.

To fit the VAR model, we need to choose the order of the model. The Akaike Information Criterion and Akaike's Final Prediction Error lead us to a model of order 2, while the Hannan-

[3] Both integrals can be numerically computed using the Cuhre algorithm of the Cuba library by Thomas Hahn. This library in R is presented by the package `R2Cuba`.

Quinn Criterion and the Schwarz-Bayes Criterion show optimality at order 1. To keep the model as simple as possible we opt for the order 1. With the number of assets $n = 2$ and the order of the VAR model $p = 1$, Equation 66 yields:

$$\begin{aligned} r_{1,t} &= m_1 + \theta_{1,1} \cdot r_{1,t-1} + \theta_{1,2} \cdot r_{2,t-1} + \epsilon_{1,t} \\ r_{2,t} &= m_2 + \theta_{2,1} \cdot r_{1,t-1} + \theta_{2,2} \cdot r_{2,t-1} + \epsilon_{2,t} \end{aligned} \tag{69}$$

Using the ordinary least squares (OLS) method, the parameters of the VAR model are estimated with corresponding values depicted in (70). We have tested the data on heteroskedasticity by White's test yielding p-value 0.0316. Even though this value indicates strong evidence against the null hypothesis of homoskedasticity, we stick to the VAR model due to its simplicity.

$$\begin{aligned} r_{1,t} &= 0.0035 + 0.0126 \cdot r_{1,t-1} - 0.0141 \cdot r_{2,t-1} + \epsilon_{1,t} \\ r_{2,t} &= 0.0064 - 0.1270 \cdot r_{1,t-1} + 0.1605 \cdot r_{2,t-1} + \epsilon_{2,t} \end{aligned} \tag{70}$$

The error terms' correlation matrix and estimates of standard deviations are shown in Table 5.3. Knowledge of these estimates makes it possible to discretize the random part of the model and build the scenario tree.

	$\epsilon_{1,t}$, bonds	$\epsilon_{2,t}$, stocks
$\epsilon_{1,t}$, bonds	1	-0.1743
$\epsilon_{2,t}$, stocks	-0.1743	1
Std. deviation, $\sigma(\epsilon_{i,t})$	0.0108	0.0440

Table 5.3: Standard deviation and correlation structure estimates for the residuals.

Example of the discretization of an error term for scenario tree generation. The error term $(\epsilon_{1,t}, \epsilon_{2,t})$ is assumed to have bivariate normal distribution. In Figure 5.3, the contour plot of the joint probability density function of such distribution is depicted. We illustrate the example of discretization with $k = 3$ points per variable (in our context the assets), which results in a total of $k^n = 3^2 = 9$ points. The dashed lines are the corresponding quantiles of the marginal distributions (the boundaries of hypercubes), and finally points represents the outcomes of the discretized equivalent. The respective numbers indicate the probabilities assigned to these points. At the first glance, it might seem that the points are symmetrical along the x- and y-axes, which together with identical probabilities would yield to the absence of correlation. However, this is resolved by probabilities of corner points that differ significantly (0.0884 for the bottom left against 0.1346 for the top left corner for instance). This method of discretization is extensible to an arbitrary number of points and limited only by the precision of the integral approximations and the available memory for the computations.

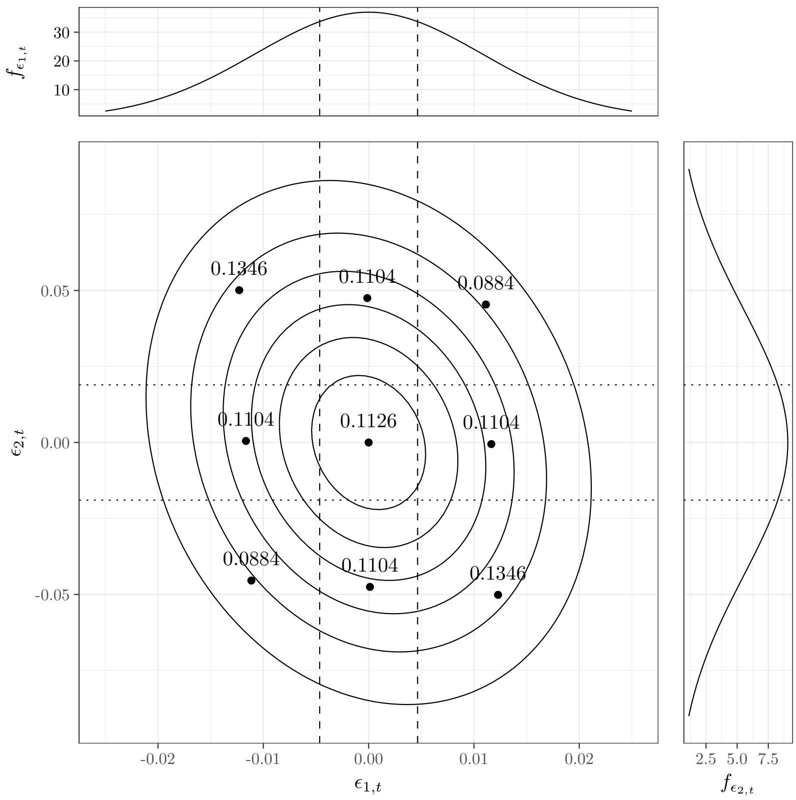

Figure 5.3: Contour plot of the probability density function of bivariate normally distributed errors with discretized counterpart and marginal densities.

5.3.2 Convergence analysis of the reference case

Intuitively it is clear that the lager the number of intervals k the better the approximation of the characteristics of the desired random variables. However, even for the rather "trivial" case that we consider, the behavior of the optimal solution with respect to k is not obvious. There are two questions of interest: does the optimal solution converge when increasing k, and, if yes, what is the adequate and sufficient value for k in practical application considering a trade-off between complexity and misspecification. Alternatively, the problem can be described as follows: With very large k the problem size becomes impractical (storage limitations, see Appendix 5.D), and on the other hand, a too small value of k implies a very large error in the optimal solution.

First, we fix the wealth parameters as described in Table 5.1. Then, we vary the value of k for different T up to the values for which we run into storage limitations (see Appendix 5.D for details). Specifically for the planning horizon $T = 2$ we allocate the problem for the number of intervals $k = 2, ..., 19$; for $T = 3$, $k = 2, ..., 7$, for $T = 4$, $k = 2, ..., 4$, for $T = 5$, $k = 2, 3$. Finally, for $T = 6, 7, 8$ we can only use $k = 2$ (we avoid the consideration of one-stage case with $T = 1$ as a trivial one). The explicit formulation of the simplest considered problem with $k = 2$ and $T = 2$ can be seen in Appendix 5.B. The values of the optimal investment in bonds at the initial time $t = 0$ for all possible cases are shown in Appendix 5.C.

We consider the case with $T = 2$ in details since for this case we can derive optimal proportions for 18 values of k ($k = 2, ..., 19$). The optimal shares to be invested in bonds are depicted in Figure 5.4 (a) as a function of k. The shape of the curve indicates convergence to the optimal investment in bonds. Since the number of possible k is still rather small, giving an estimate of the value to which the function converges is difficult. Thus, we assume the optimal amount invested in bonds for $k = 19$ to be the exact value, and based on this value we calculate relative errors for smaller k. These relative errors are shown in Figure 5.4 (b) and also support our hypothesis of convergence. Furthermore, for k larger than eight, the absolute value of relative errors is less than 0.5%, which is a good indication that the true value is located near the assumed one. Also, from this plot we can observe that using only $k = 2$ points per variable would lead to an error of around 5% of the optimal value.

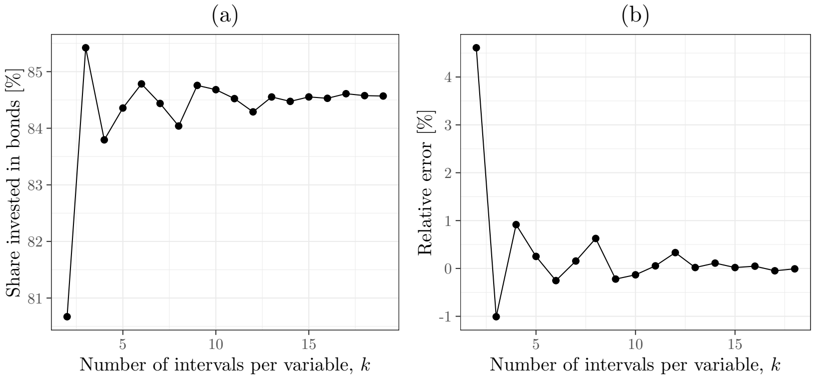

Figure 5.4: Share invested in bonds (a) and relative error (b) as a function of the number of intervals per variable k on a time horizon $T = 2$.

Figure 5.5 depicts both the probability of the surplus (a) and the expected value of surplus given excess (b). Both parts support the hypothesis of convergence, and it seems that both functions converge. Also, it can be noted that a variability decreases significantly starting from $k = 6$ for both variables. The plot indicates that in around 70% of the cases the value of assets at $T = 2$ is greater than target wealth, and given this objective is attained, the surplus on average is around 18 money units (1.8 % of initial investment).

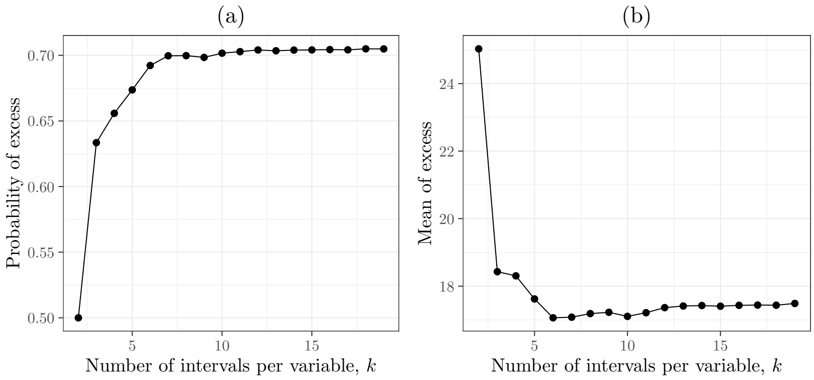

Figure 5.5: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the number of intervals per variable k on a time horizon $T = 2$.

It is interesting to see that the probability of deficit and the mean of the deficit behave slightly differently when comparing to the excess in Figure 5.5. The value of the mean deficit stabilizes starting from $k = 12$. Given the strict signs in Equations 63 and 65, we do not include the cases when the value of assets exactly equals to the target wealth ($A_T - L_T = 0$) into the estimation of both expected deficit and expected excess. The explanation might be that after $k = 12$ these zeros become very small negative values which are included in the computation and, therefore, bias the estimate.

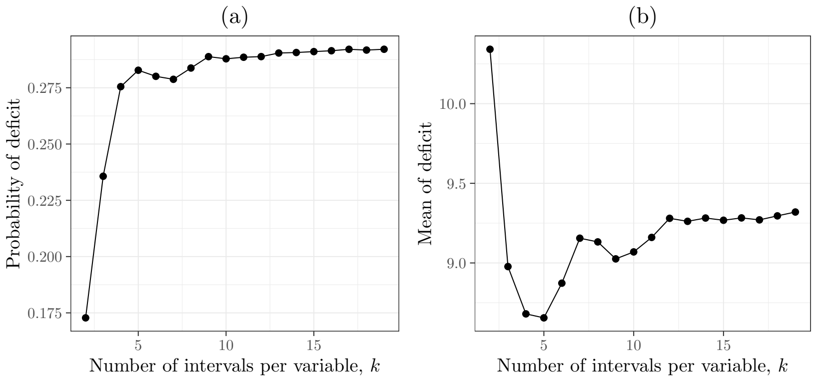

Figure 5.6: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the number of intervals per variable k on a time horizon $T = 2$.

To summarize the above points, we observe that values for k larger than 12 seem to give stable estimates of the desired parameters with minimal error in our example. However, such recommendation would be hardly implementable due to the size of the problem. Thus, we suggest using at least four intervals per variable as compromise between precision and complexity with focus on the initial asset allocation.

5.3.3 Sensitivity analyses

In this section, we present selected sensitivity analyses on the previously discussed problem: we vary the planning horizon T, the target wealth L_2, the shortage penalty r, the bonds mean return m_1, and the volatility of stocks residuals $\sigma_{t,2}$

Planning horizon T. First, we are interested in the effect of extending the planning horizon T. For analyzing the sensitivity of key values to changes in the planning horizon T we use the same values of the utility function, the target wealth, the initial investment and returns model. We fix k at two, which allows us to store problems with larger T. Though, the same conclusions can be extended and applied to higher values of k.

The percentages invested in bonds at time $t = 0$ for different T are shown in Figure 5.7. The function has a linear shape with slight deviations: the value of the share decreases on average by 4% with one-month increase of T. This means that the institution invests more aggressively and risky (i.e. with higher percentage in stocks) for higher T. This fact goes along with natural intuition: the average interest (which is positive) accumulated over the longer T is larger and even nodes with small interest when combined together produce higher return. Long investment horizon permits to better handle the stock volatility (invest more in stocks). This triggers the shortfall to appear less frequently. Thus, with the same level of utility function the investor would invest more into volatile asset class (stocks).

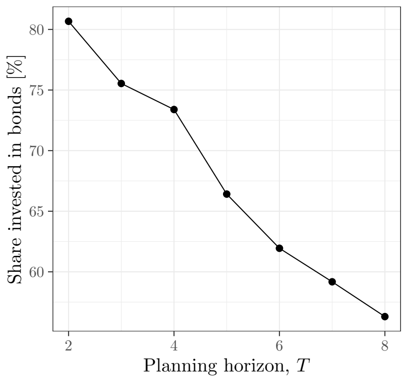

Figure 5.7: Share invested in bonds as a function of the planning horizon T for a number of intervals per variable $k = 2$.

Further, the probability of surplus as well as the mean of the excess increase in T (Figure 5.8). The explanation of this increase coincides with the explanation of decreases in bonds' shares: longer planning horizons T and lower bonds investments induces higher average accumulated interests.

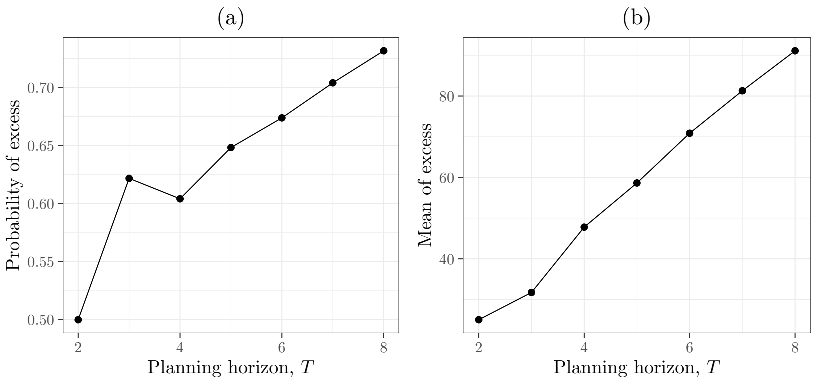

Figure 5.8: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the planning horizon T for number of intervals per variable $k = 2$.

It is interesting to see from Figure 5.9 that while the probability of deficit decreases with T, the mean of a such deficit increases. It means that missing the target wealth would happen less often, but if it happens, the value of mismatch is larger. This is probably linked to the fact that the investor invests heavily in stocks with increase of T. Higher share invested in stocks yield the higher volatility, i.e. higher absolute values of both positive and negative returns, which also explains the increase in the average surplus.

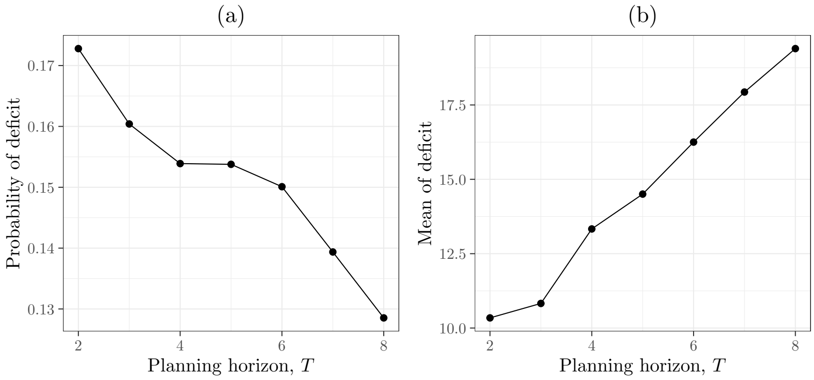

Figure 5.9: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the planning horizon T for number of intervals per variable $k = 2$.

Target wealth L_2. For studying the impact of changes in the target wealth L_T, the number of intervals k and the planning horizon T are set at levels eight and two, respectively. Then, we vary L_2 in the set of values $\{1000, 1010, 1020, ..., 1120\}$.

Figure 5.10 supports the intuition: higher values of target wealth can only be attained if the pension fund invests larger proportion in stocks. Indeed, after a small interval of increase $(1000, 1010)$ the function is rather flat until 1040, and then the function starts to decrease rapidly. Around a value $L_2 = 1100$ the initial wealth is fully invested in stocks. The interval until the rapid decrease is characterized by the feasibility to attain without investing into stocks. High values of L_2, in contrast, require investment in stocks.

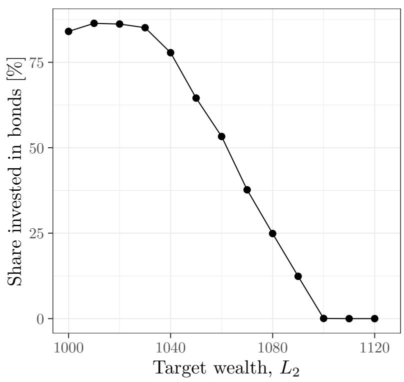

Figure 5.10: Share invested in bonds as a function of the target wealth L_2.

The shape of the function that represents the probability of excess (part (a) of Figure 5.11) is strictly decreasing. This function changes its shape at $L_2 = 1040$. This rapid change corresponds to the curve of the bonds' share in Figure 5.10. At the first glance, it seems that the means of excess (part b) has an anomalous shape. However, considering this function with care, all parts can be explained. For instance, the parabola in the interval $(1000, 1040)$ is related to the flat region of the shares in bonds. Then we see again a rather smooth segment up to the value 1100. This value is characterized as a point at which the investor allocates all wealth in stocks. Further, the increase of this function is related to the fact that for larger target wealth L_2 the surplus occurs in fewer nodes. While the number of such nodes decreases, we are left with nodes of extreme returns, which drive the final wealth.

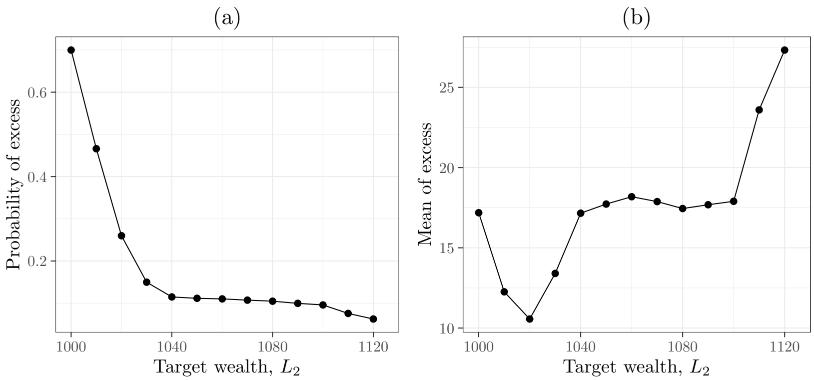

Figure 5.11: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the target wealth L_2.

Both functions of Figure 5.12 related to the deficit are increasing. The larger the target wealth the higher the chances of deficits. At the same time, the average value of this mismatch increases.

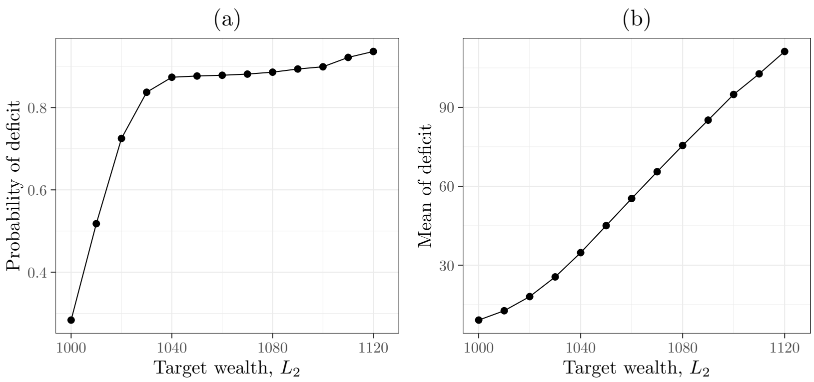

Figure 5.12: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the target wealth L_2.

Shortage penalty r. By varying the shortage penalty r from one to five with step 0.25 we study the sensitivity to the changes of the utility function. The shares invested in bonds in Figure 5.13 grows with r. Higher shortage penalty indicates a stronger risk aversion of investors. Therefore, the increase seems to be consistent with the concept of a utility function. We outline that the proportion invested in bonds is at level 0% not only for the case $r = 1$, but also for $r = 1.25$.

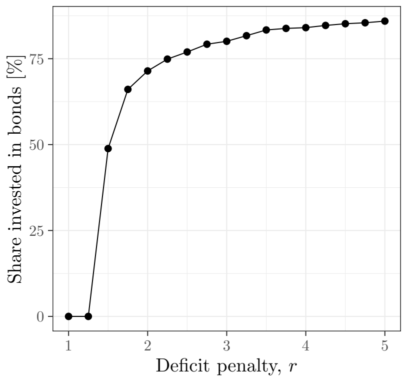

Figure 5.13: Share invested in bonds as a function of the shortage penalty r.

Looking at both curves in Figure 5.14 it is possible to capture the intuition of the trade-off between the probability of excess and its average size. In other words, the driving characteristic for risky investors is the average return, while risk averse investors tend to increase the probability to have a surplus, disregarding how big the surplus would be.

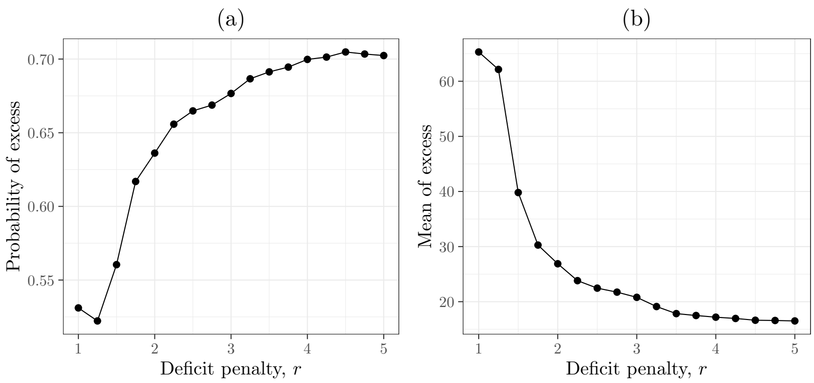

Figure 5.14: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the shortage penalty r.

Given the increase of the probability of surplus, the decrease in the probability of deficit is clear. The shortage penalty directly influences the value of the average deficit, that is, the increase in shortage penalty is linked to a decrease in the average deficit. This is illustrated in Figure 5.15.

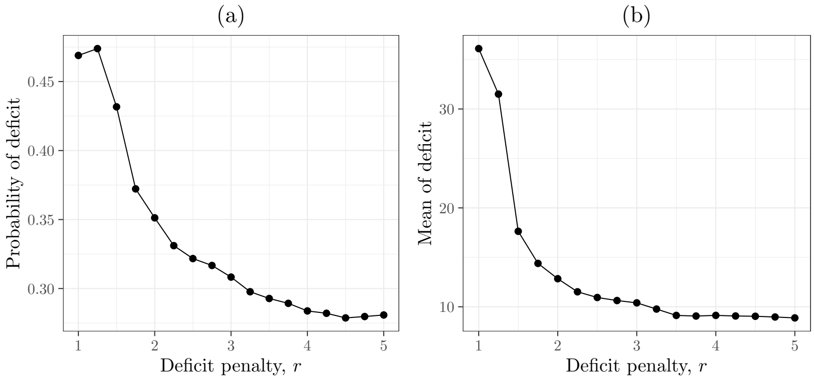

Figure 5.15: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the shortage penalty r.

Bonds' mean return m_1. To study the impact of changes in the average return of bonds, we vary m_1 (the mean return of bonds, see Equation 69 for details) in the range from -0.008 to 0.008 by step of 0.001 (recall the original estimate is 0.0034). This is particularly of interest in the current low risk-free rate environment found in Switzerland. This analysis helps in understanding how the optimal asset allocation is misspecified, if the estimate of bond returns is biased. The curve in Figure 5.16 is strictly increasing and concave. In addition, we highlight that even for negative values of average bond returns, the optimal allocation still includes shares in bonds. This effect is explained by the consequence of the diversification. Indeed, stocks and bonds are negatively correlated in our case (cf. Table 5.2). Furthermore, a negative average return does not imply that returns are negative in all nodes.

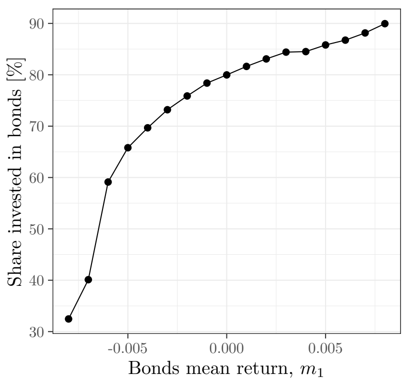

Figure 5.16: Share invested in bonds as a function of the bonds' mean return m_1.

Considering the probability of excess in Figure 5.17 (a), it should not be left unmentioned that the function decreases for decreasing returns until $m_1 = -0.004$. Nevertheless, beyond this value (going more negative) the function is increasing. The decrease in the mean excess is due to the fact, that the fund invests more in bonds when average bond returns increase, which induces lower average returns.

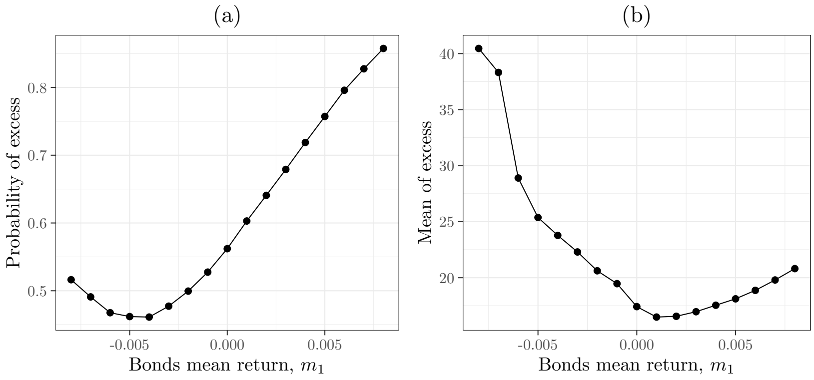

Figure 5.17: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the bonds' mean return m_1.

The dominating decrease of the probability of deficit coincides with the shape of the probability of excess curve (Figure 5.18 a). In addition, the decrease of the average deficit can be explained by the increase of bond shares.

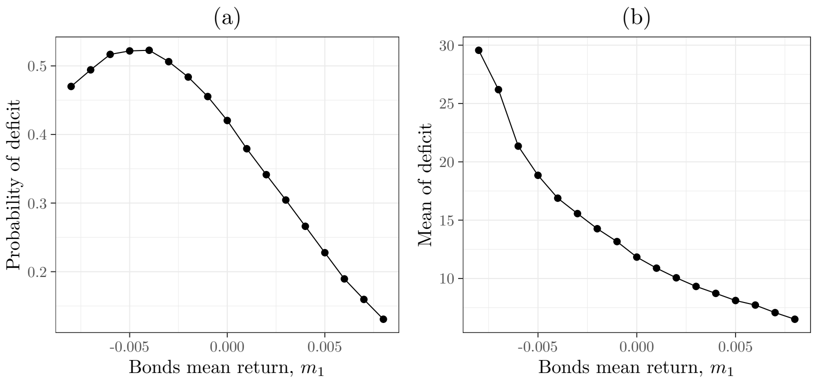

Figure 5.18: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the bonds' mean return m_1.

Volatility of stocks' residuals $\sigma_{t,2}$. The last we study is the standard deviation of stock's residuals $\sigma_{t,2}$. This value represents the volatility of stocks in our context and is often difficult to estimate. By utilizing a VAR model, we explicitly assume the stationarity of the return time series. Thus, the variance remains the same through the time points, and its value plays a crucial role in the optimal solution.

We concentrate on optimal solutions for models with different $\sigma_{t,2}$: we form the sequence of values in the range from 0.01 to 0.05 with step size 0.003 (the originally estimated reference value is 0.0438). In this range the share invested in bonds increases, which means that, for the same utility function with greater stocks' volatility, investors invest a larger proportion in bonds.

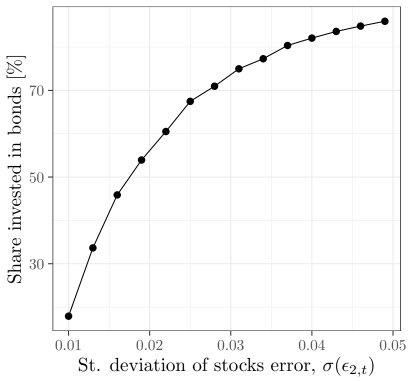

Figure 5.19: Share invested in bonds as a function of the volatility of stocks' residuals $\sigma_{t,2}$.

The probability of excess (Figure 5.20 a) smoothly decreases with respect to the standard deviation of stock's error. At the same time, the mean excess (Figure 5.20 b) behaves differently depending on the interval under consideration. The maximum values are obtained in a range of 0.02 to 0.03. The rapid increase preceding this range means that investors who are concerned about the mean excess should consider more volatile stocks, with standard deviation in the range $(0.02, 0.03)$. Starting from 0.03 the mean excess slowly decreases.

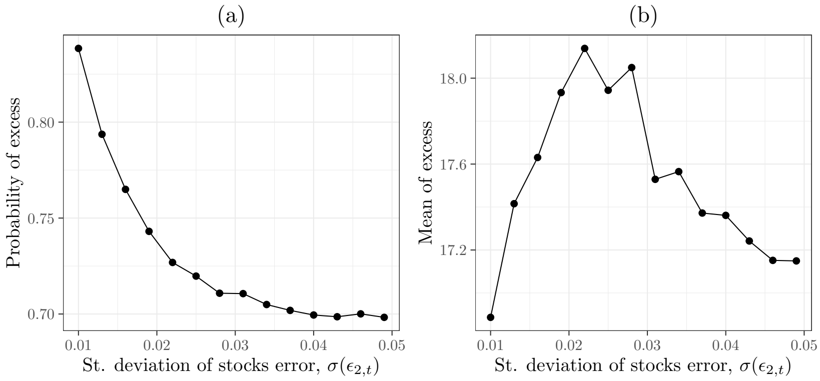

Figure 5.20: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the volatility of stocks' residuals $\sigma_{t,2}$.

As expected, the probability of deficit increases symmetrically as the probability of excess decreases. A similar increase can be found for the mean of the deficit. Figure 5.21 presents both plots.

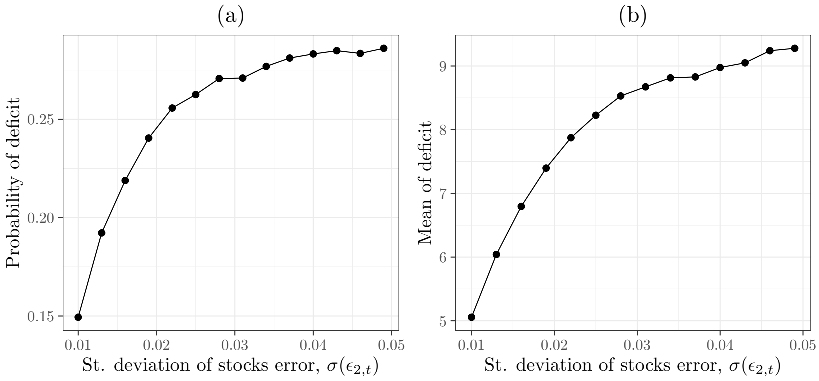

Figure 5.21: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the volatility of stocks' residuals $\sigma_{t,2}$.

5.3.4 Comparison with Monte Carlo simulation

Stepping away from SP, we consider the other common approach for determining asset allocations, namely the Monte Carlo method. In order to compare both methods we must choose same reference framework, and therefore, we use the time horizon $T = 2$ with the same returns data and model described in Section 5.3. The buy & hold strategy (see Perold and Sharpe, 1988) is used, that is, once assets are bought at stage $t = 0$ they are not reallocated at the intermediate points. A sketch of an algorithm is as follows:

1. Simulate $N = 10000$ paths of the VAR model
2. Fix the initial asset allocation at $t = 0$
3. Using buy & hold strategy to calculate the final wealth for each of the simulated paths
4. According to Equations 62-65, estimate the quantities of interest

This algorithm is applied to the set of different initial asset allocations. As long as only two asset classes are considered, we vary the share invested in bonds from 0% to 100% (with steps of 1%). In order to keep lines smooth we use the same seed for the various initial asset allocations.

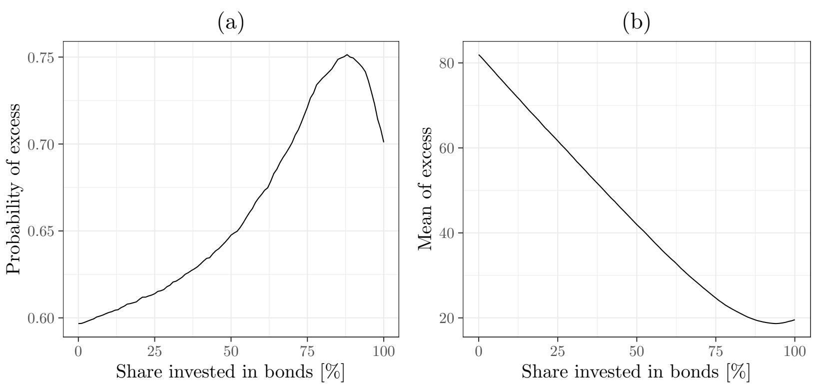

Figure 5.22: Probability of excess (a) and expected value of the excess given surplus (b) as a function of the shares invested in bonds for Monte Carlo simulations.

In Figures 5.22 and 5.23, we report the probability of excess, the mean of surplus given excess, the probability of deficit and the mean of shortage given deficit as functions of the share invested in bonds. The probability of excess increases until a maximum is reached at 88%, and then declines. This level of 88% does not coincide with the optimal value found with stochastic optimization (about 84-85%), even though it is relatively close. However, these values cannot be compared, as long as 88% is obtained by maximizing only the probability of excess, while in SP we use the utility function concept. The probability of deficit has symmetrical shape, i.e. decreases and then ascents after a minimum. We also see that both the mean of excess and the mean of deficit are decreasing functions. Increasing the share invested in bonds effectively decreases the volatility of the whole portfolio.

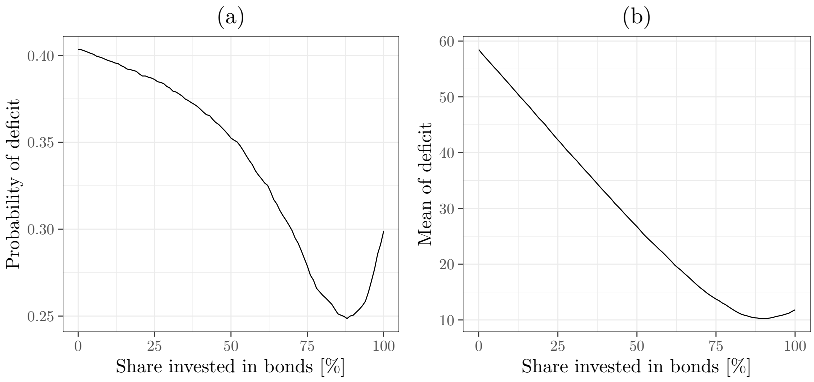

Figure 5.23: Probability of deficit (a) and expected value of the deficit given shortage (b) as a function of the shares invested in bonds for Monte Carlo simulations.

This approach does not answer the question what initial allocation is optimal. However, it allows seizing the sensitivity and the dependence of the allocation and the final portfolio value on the chosen parameters.

5.4 Concluding remarks

In this note, we investigate the properties of stochastic programming applied to the asset allocation in a pension fund. By means of a simple model, we examine the convergence of key characteristics with respect to the quality of the scenario tree, the sensitivity to model parameters, and the effect of the estimations of the economic parameters. Our main observations can be summarized as follows:

Convergence with respect to the discretization parameter. In order to construct the scenario tree, the underlying stochastic process of assets must be discretized. We employ the "bracket-mean" method, for which the number of marginal intervals parametrizes the precision of the discrete approximation. Our findings empirically prove the convergence of the initial asset allocation, the probability of the deficit and the expected value of the deficit with respect to increases of the precision.

Sensitivity analysis to changes in internal parameters. The objective of our stochastic programming problem is to maximize the expected value of the utility function, which is parameterized by the deficit penalty and the target wealth. Further, the objective function is estimated at a certain point in the future, defined as the planning horizon. Expectedly, our main observations are as follows: (1) the initial share invested in bonds increases with the shortage penalty, decreases with the target wealth, and decreases with the target horizon; (2) the probability of deficit decreases with the shortage penalty, increases with the target wealth, and decreases with the target horizon; (3) the expected value of deficit given a shortage decreases with the shortage penalty, increases with the target wealth, and increases with the target horizon.

Effect of misestimation of the econometric model parameters. The scenario tree is sensitive to estimation errors. To deepen our understanding, we vary the expected bond return and the volatility of stocks. We show that the initial share invested in bonds increases with the mean bond return. It is interesting to see that even for negative values of the expected return of bonds the share invested stays non-zero. A similarly increasing curve can be observed when the volatility of stocks increases.

Bibliography

Bertsimas, D. and J. N. Tsitsiklis, 1997, *Introduction to Linear Optimization.* Athena Scientific Belmont, MA.

Birge, J. R. and F. Louveaux, 2011, *Introduction to Stochastic Programming.* Springer Science & Business Media.

Bogentoft, E., H. E. Romeijn, and S. Uryasev, 2001, Asset/Liability Management for Pension Funds Using CVaR Constraints, *The Journal of Risk Finance*, 3(1):57–71.

Bütler, M. and M. Ruesch, 2007, Annuities in Switzerland, *Policy Research Wroking Paper, The World Bank.*

Cariño, D. R., T. Kent, D. H. Myers, C. Stacy, M. Sylvanus, A. L. Turner, K. Watanabe, and W. T. Ziemba, 1994, The Russell-Yasuda Kasai Model: An Asset/Liability Model for a Japanese Insurance Company Using Multistage Stochastic Programming, *Interfaces*, 24(1):29–49.

Cariño, D. R., D. H. Myers, and W. T. Ziemba, 1998, Concepts, Technical Issues, and Uses of the Russell-Yasuda Kasai Financial Planning Model, *Operations Research*, 46(4):450–462.

Cariño, D. R. and W. T. Ziemba, 1998, Formulation of the Russell-Yasuda Kasai Financial Planning Model, *Operations Research*, 46(4):433–449.

Dert, C. L., 1995, *Asset Liability Management for Pension Funds. A Multistage Chance Constrained Programming Approach*, Ph.D. thesis, Erasmus University Rotterdam.

Dondi, G. A., 2005, *Models and Dynamic Optimisation for the Asset and Liability Management of Pension Funds*, Ph.D. thesis, Swiss Federal Institute of Technology in Zurich.

Drijver, S. J., 2005, *Asset Liability Management for Pension Funds Using Multistage Mixed-Integer Stochastic Programming*, Ph.D. thesis, University of Groningen.

Dupacová, J., G. Consigli, and S. W. Wallace, 2000, Scenarios for Multistage Stochastic Programs, *Annals of Operations Research*, 100(1):25–53.

Fleten, S. E., K. Høyland, and S. W. Wallace, 2002, The Performance of Stochastic Dynamic and Fixed Mix Portfolio Models, *European Journal of Operational Research*, 140(1):37–49.

Geyer, A. and W. T. Ziemba, 2008, The Innovest Austrian Pension Fund Financial Planning Model InnoALM, *Operations Research*, 56(4):797–810.

Heitsch, H. and W. Römisch, 2009, Scenario Tree Modeling for Multistage Stochastic Programs, *Mathematical Programming*, 118(2):371–406.

Hibiki, N., 2006, Multi-Period Stochastic Optimization Models for Dynamic Asset Allocation, *Journal of Banking & Finance*, 30(2):365–390.

Hilli, P., M. Koivu, T. Pennanen, and A. Ranne, 2007, A Stochastic Programming Model for Asset Liability Management of a Finnish Pension Company, *Annals of Operations Research*, 152(1):115–139.

Hochreiter, R. and G. C. Pflug, 2007, Financial Scenario Generation for Stochastic Multi-Stage Decision Processes as Facility Location Problems, *Annals of Operations Research*, 152(1):257–272.

Høyland, K. and S. W. Wallace, 2001, Generating Scenario Trees for Multistage Decision Problems, *Management Science*, 47(2):295–307.

Kallberg, J. G. and W. T. Ziemba, 1983, Comparison of Alternative Utility Functions in Portfolio Selection Problems, *Management Science*, 29(11):1257–1276.

Kaut, M. and S. W. Wallace, 2007, Evaluation of Scenario-Generation Methods for Stochastic Programming, *Pacific Journal of Optimization*, 3(2):257–271.

Kaut, M. and S. W. Wallace, 2011, Shape-Based Scenario Generation Using Copulas, *Computational Management Science*, 8(1-2):181–199.

Keefer, D. L., 1994, Certainty Equivalents for Three-Point Discrete-Distribution Approximations, *Management Science*, 40(6):760–773.

Keefer, D. L. and S. E. Bodily, 1983, Three-Point Approximations for Continuous Random Variables, *Management Science*, 29(5):595–609.

Koivu, M., T. Pennanen, and A. Ranne, 2005, Modeling Assets and Liabilities of a Finnish Pension Insurance Company: a VEqC Approach, *Scandinavian Actuarial Journal*, 2005(1):46–76.

Kouwenberg, R., 2001, Scenario Generation and Stochastic Programming Models for Asset Liability Management, *European Journal of Operational Research*, 134(2):279–292.

Kusy, M. I. and W. T. Ziemba, 1986, A Bank Asset and Liability Management Model, *Operations Research*, 34(3):356–376.

Markowitz, H., 1952, Portfolio Selection, *The Journal of Finance*, 7(1):77–91.

Miller III, A. C. and T. R. Rice, 1983, Discrete Approximations of Probability Distributions, *Management Science*, 29(3):352–362.

OECD, 2015, *Pension markets in focus*.

Pearson, E. S. and J. W. Tukey, 1965, Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves, *Biometrika*, 52(3/4):533–546.

Pennanen, T. and M. Koivu, 2002, Integration Quadratures in Discretization of Stochastic Programs, *Stochastic Programming E-print Series*, 11.

Pennanen, T. and M. Koivu, 2005, Epi-Convergent Discretization of Stochastic Programs via Integration Quadratures, *Numerische Mathematik*, 100(1):141–163.

Perold, A. F. and W. F. Sharpe, 1988, Dynamic Strategies for Asset Allocation, *Financial Analysts Journal*, 44(1):16–27.

Pflug, G., 2001, Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization, *Mathematical Programming*, 89(2):251–271.

Reynisson, G. M., 2012, *Asset Liability Management for Icelandic Pension Funds*, Ph.D. thesis, University of Iceland.

Shapiro, A., D. Dentcheva, and A. Ruszczynski, 2009, *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.

Smith, J. E., 1993, Moment Methods for Decision Analysis, *Management Science*, 39(3):340–358.

Swiss Federal Statistical Office, 2017, *Continuous decline in defined benefit plans.*

Toukourou, Y. A. F., 2008, *Asset Liability Management and Joint Mortality Modelling in Old Age Insurance*, Ph.D. thesis, University of Lausanne.

Toukourou, Y. A. F. and F. Dufresne, 2018, On Integrated Chance Constraints in ALM for Pension Funds, *ASTIN Bulletin: The Journal of the IAA*, 48(2):571–609.

Willis Tower Watson, 2016, *Pension Risk Study on the funding ratio of SLI companies.*

5.A Appendix: Scenario tree generation techniques

Sampling methods. The sampling method is relatively straightforward: for a VAR model, given a previous historical realization the distribution of the vector is known (the random part is only the error term), and a pseudo random vector drawn from its distribution can be generated. This routine is repeated a certain number of times (the number of simulations), and conditional on the obtained values the same procedure is performed going further in time. Several methods of variance reduction can be applied. For instance, the antithetic method allows exploiting the symmetry of some distribution families.

However, in order to reflect the main properties of the distribution the number of simulations must be sufficiently large, which challenges the optimization problem. The size and complexity of the scenario tree grows as a polynomial of order four increases in the length of the sample. Furthermore, the solutions will differ depending on the realizations. Finally, the following trivial example can illustrate the weakness of using a sampling method for heavy tailed distributions. Assume a discrete distribution of returns that have an extremely negative value with occurrence probability $\frac{1}{1000}$. In order to incorporate this value in the solution one needs on average of 1000 simulations, and still there is no guarantee that this value will be generated. At the same time in order to ensure a "uniform" and "even" spread of points one can use quasi-random sequences, i.e. Halton or Sobol sequences to model uniform distributions and then apply inversion of these points for desired distribution.

"Bracket-mean" extension. It is worth mentioning another possible approach for bivariate normally-distributed vectors ($n = 2$). First, one discretizes one of the variables, say $\epsilon_{1,t}$ into k points. Then, given $\epsilon_{1,t}$, the conditional distribution of $\epsilon_{2,t}$ is known:

$$\epsilon_{2,t}|\epsilon_{1,t} = x \sim N(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x - \mu_1), \quad (1 - \rho^2)\sigma_2^2), \tag{71}$$

where μ_1, σ_1 and μ_2, σ_2 are the corresponding expected value and standard deviation parameters of $\epsilon_{1,t}$ and $\epsilon_{2,t}$; ρ is the correlation coefficient between $\epsilon_{1,t}$ and $\epsilon_{2,t}$. Knowing the conditional distribution of $\epsilon_{2,t}$ allows us to discretize $\epsilon_{2,t}$ for each point $\epsilon_{1,t} = x$, which results in desired k^2 discrete points. This method is technically more trivial and tractable, because it only requires the calculation of univariate integrals in contrast to "bracket-mean" that assumes bivariate ones. This method is limited to bivariate random vectors. In addition, while $\epsilon_{2,t}$ can take k^2 distinct values, $\epsilon_{1,t}$ has only k.

Moment matching method via integration quadratures. The undesirable underestimation of moments in the "bracket-mean" method can be avoided by utilizing Gaussian quadrature, also known as moment matching method, firstly proposed by Miller III and Rice (1983) and discussed by Smith (1993). However, this approach uses too many higher-order moments. This makes this method hardly applicable in practice, for which typically only stable and robust moments' estimates are available. In addition, Smith (1993) warns that this method tends to "chase the tail", meaning that for instance for three-point discretizations the upper point could be higher than the 99-% quantile. Further development of this method can be found in Pennanen and Koivu (2002) and Pennanen and Koivu (2005). Finally, Høyland and Wallace (2001) propose to minimize the square difference (as a measure of distance) between the true distribution statistical properties and its discrete version properties. The latter method is more flexible, because it allows an expert to specify statistical properties which are relevant depending on a given problem.

"Optimal discretization". Finally, the method proposed by Pflug (2001), see also Hochreiter and Pflug (2007), and discussed by Kaut and Wallace (2007) is reviewed below. The idea of their optimal discretization is to minimize the so-called approximation error, that is the difference between the optimal value of the objective function of the original problem and the problem that utilizes the approximated distribution. Therefore, the whole scenario tree is built at once. Kaut and Wallace (2007) report, that the model performs better than the above-mentioned models. However, the model requires solving a non-linear optimization problem. With increasing of the size of the scenario tree, the model becomes impractical to solve in the sense of computational techniques.

This list of scenario generating techniques has not the aim of being comprehensive and exhaustive, other techniques exist. For instance, Keefer and Bodily (1983) and Keefer (1994) consider extensions of three-point approximations proposed by Pearson and Tukey (1965) including multivariate cases. A more systematic approach of comparison of some scenario tree generation methods can be found, for example in Kaut and Wallace (2007) and Kouwenberg (2001).

5.B Appendix: Example of a stochastic problem

Based on data in Section 5.3.1 the scenario tree is built and depicted on Figure 5.24 with corresponding assets' returns and probabilities of a scenario's path to occur. We use parameter values as laid out in Table 5.1.

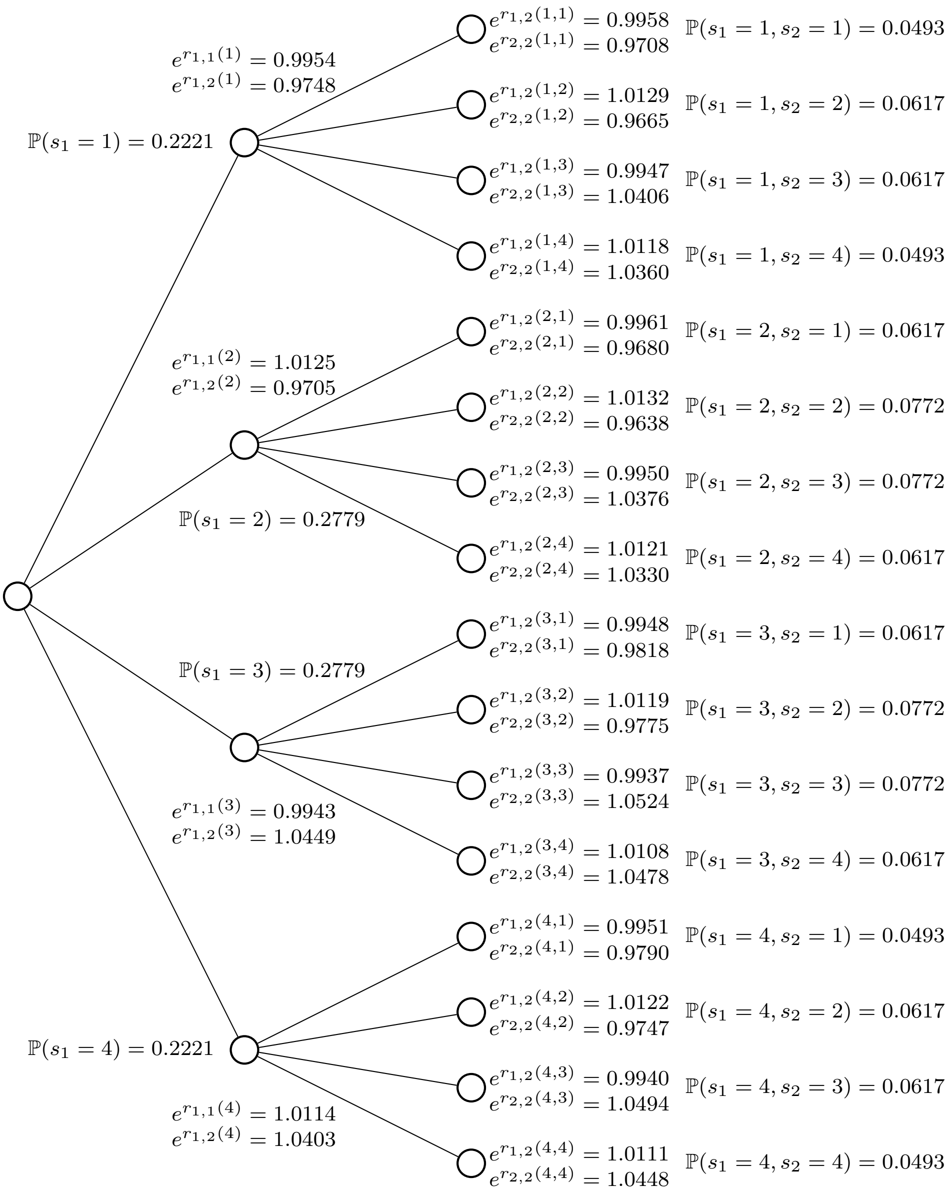

Figure 5.24: Illustration of a scenario tree with $T = 2$ stages, $n = 2$ asset classes and $k = 2$ intervals per asset.

$$
\begin{aligned}
\max \quad & 0.0493 \cdot (y(1,1) - 4 \cdot w(1,1)) \\
& 0.0617 \cdot (y(1,2) - 4 \cdot w(1,2)) + \\
& 0.0617 \cdot (y(1,3) - 4 \cdot w(1,3)) + \\
& 0.0493 \cdot (y(1,4) - 4 \cdot w(1,4)) + \\
& 0.0617 \cdot (y(2,1) - 4 \cdot w(2,1)) + \\
& 0.0772 \cdot (y(2,2) - 4 \cdot w(2,2)) + \\
& 0.0772 \cdot (y(2,3) - 4 \cdot w(2,3)) + \\
& 0.0617 \cdot (y(2,4) - 4 \cdot w(2,4)) + \\
& 0.0617 \cdot (y(3,1) - 4 \cdot w(3,1)) + \\
& 0.0772 \cdot (y(3,2) - 4 \cdot w(3,2)) + \\
& 0.0772 \cdot (y(3,3) - 4 \cdot w(3,3)) + \\
& 0.0617 \cdot (y(3,4) - 4 \cdot w(3,4)) + \\
& 0.0493 \cdot (y(4,1) - 4 \cdot w(4,1)) + \\
& 0.0617 \cdot (y(4,2) - 4 \cdot w(4,2)) + \\
& 0.0617 \cdot (y(4,3) - 4 \cdot w(4,3)) + \\
& 0.0493 \cdot (y(4,4) - 4 \cdot w(4,4))
\end{aligned}
\tag{72}
$$

$$
\begin{gathered}
\text{subject to:} \quad x_{1,0} + x_{2,0} = 1\,000 \\
0.9954 \cdot x_{1,0} + 0.9748 \cdot x_{2,0} - x_{1,1}(1) - x_{2,1}(1) = 0 \\
1.0125 \cdot x_{1,0} + 0.9705 \cdot x_{2,0} - x_{1,1}(2) - x_{2,1}(2) = 0 \\
0.9943 \cdot x_{1,0} + 1.0449 \cdot x_{2,0} - x_{1,1}(3) - x_{2,1}(3) = 0 \\
1.0114 \cdot x_{1,0} + 1.0403 \cdot x_{2,0} - x_{1,1}(4) - x_{2,1}(4) = 0 \\
0.9958 \cdot x_{1,1}(1) + 0.9708 \cdot x_{2,1}(1) - y(1,1) + w(1,1) = 1\,000 \\
1.0129 \cdot x_{1,1}(1) + 0.9665 \cdot x_{2,1}(1) - y(1,2) + w(1,2) = 1\,000 \\
0.9947 \cdot x_{1,1}(1) + 1.0406 \cdot x_{2,1}(1) - y(1,3) + w(1,3) = 1\,000 \\
1.0118 \cdot x_{1,1}(1) + 1.0360 \cdot x_{2,1}(1) - y(1,4) + w(1,4) = 1\,000 \\
0.9961 \cdot x_{1,1}(2) + 0.9680 \cdot x_{2,1}(2) - y(2,1) + w(2,1) = 1\,000 \\
1.0132 \cdot x_{1,1}(2) + 0.9638 \cdot x_{2,1}(2) - y(2,2) + w(2,2) = 1\,000 \\
0.9950 \cdot x_{1,1}(2) + 1.0376 \cdot x_{2,1}(2) - y(2,3) + w(2,3) = 1\,000 \\
1.0121 \cdot x_{1,1}(2) + 1.0330 \cdot x_{2,1}(2) - y(2,4) + w(2,4) = 1\,000 \\
0.9948 \cdot x_{1,1}(3) + 0.9818 \cdot x_{2,1}(3) - y(3,1) + w(3,1) = 1\,000 \\
1.0119 \cdot x_{1,1}(3) + 0.9775 \cdot x_{2,1}(3) - y(3,2) + w(3,2) = 1\,000 \\
0.9937 \cdot x_{1,1}(3) + 1.0524 \cdot x_{2,1}(3) - y(3,3) + w(3,3) = 1\,000 \\
1.0108 \cdot x_{1,1}(3) + 1.0478 \cdot x_{2,1}(3) - y(3,4) + w(3,4) = 1\,000 \\
0.9951 \cdot x_{1,1}(4) + 0.9790 \cdot x_{2,1}(4) - y(4,1) + w(4,1) = 1\,000 \\
1.0122 \cdot x_{1,1}(4) + 0.9747 \cdot x_{2,1}(4) - y(4,2) + w(4,2) = 1\,000 \\
0.9940 \cdot x_{1,1}(4) + 1.0494 \cdot x_{2,1}(4) - y(4,3) + w(4,3) = 1\,000 \\
1.0111 \cdot x_{1,1}(4) + 1.0448 \cdot x_{2,1}(4) - y(4,4) + w(4,4) = 1\,000 \\
x \geq 0, y \geq 0, w \geq 0
\end{gathered}
\tag{73}
$$

5.C Appendix: Investment proportions for different planning horizons

Number of intervals	Planning horizon, T						
per variable, k	2	3	4	5	6	7	8
2	80.67	75.54	73.39	66.41	61.94	59.18	56.31
3	85.42	77.57	73.55	68.51			
4	83.79	77.97	72.94				
5	84.36	77.86					
6	84.78	78.03					
7	84.44	78.28					
8	84.04						
9	84.76						
10	84.68						
11	84.52						
12	84.29						
13	84.55						
14	84.47						
15	84.55						
16	84.53						
17	84.61						
18	84.58						
19	84.57						

Table 5.4: Proportions invested in bonds for different numbers of intervals per variable and planning horizons.

5.D Appendix: Implementation details and software choice

Stochastic programming problems are demanding in terms of computational power. Building a complete SP model almost always consists of trial and errors and continuous search for better software and solutions. We carry out all experiments on two standard machines, namely: a laptop Apple MacBook Pro with 8 Gb memory, 121 Gb available storage and dual-core Intel Core i5 2.5 GHz processor (OS X 10.8.5) and a desktop with 16 Gb memory, 1 Tb storage, and octa-core Intel Core i7 3.6 GHz (Ubuntu 14.04.5).

As of today, the open-source programming language R is considered as a default tool for statistical problems. In addition, most of the recent and state-of-the-art statistical models are introduced as R packages. Finally, R is considered as a good trade-off between the poorer performance of Microsoft Excel and the complexity of multi-purposes more advanced languages in industrial applications. Thus, R was chosen as a primary software for our analysis.

Given R is tailored to statistics, analyzing financial series and fitting VAR model to data are executed smoothly and takes reasonable amount of time (seconds). We use, for example, the packages `vars` for VAR modeling and `het.test` for White's test for heteroskedasticity.

However, challenges arise before the actual optimization. Generating scenario trees with "bracket-mean" methods requires to solve or estimate multivariate integrals. The built-in function `integrate` works only for one-dimensional functions, and thus cannot be applied. The package `cubature` allows for multi-dimensional integration, which is an R wrapper around the `cubature` C library by Steven G. Johnson. However, without taking care of the integration error, the calculation of a central point (i.e. (0, 0) in case of a bivariate normal distribution) leads to a malfunction of R for a desired precision. Furthermore, estimating corner points with small probabilities leads to enormous values (as long as values of integrals must be divided over small probabilities, see Equation 68). Taking these points to account, the package `R2Cuba` is used instead, which is an R interface to the Cuba library by Thomas Hahn. This package appears to be more stable and faster compared to `cubature` (though no formal tests were conducted, which is out of scope of this paper). We utilize the deterministic algorithm `Cuhre`.

Using parameters of the VAR model we build the scenario trees, which is represented by a matrix. Each entry of this matrix represents returns of assets at a particular node. Based on the scenario tree matrix one can formulate the deterministic approximated equivalent linear problem.

There are numbers of available solvers for the linear optimization, both free open-source (e.g. `lp_solve`, `GLPK` etc.) and proprietary (`CPLEX`, `Gurobi` etc.). Though commercial solutions mostly allow for free academic use, we focus on free open-source to attain the maximum of reproducibility. Optimizers typically have their own file formats and modeling languages, such as, for instance, GMPL (MathProg) algebraic modeling language, which is compatible with the GLPK solver. At the same time, R provides several packages as interfaces for solvers.

It is also worth to mention a stand-alone package `linprog`, which is not dependent on any solvers and written entirely in R. The use of this package is very limited due to its slow speed; thus this package was used only for the smaller trivial examples. The package `linprog` requires providing a cost vector c, a constraints matrix A and a right-hand side vector b. Thus, it should be possible to allocate the matrix A in an R environment. One particularity of R is that R stores all objects in memory, which significantly limits the application in linear programming.

We attempt to use the package `lpSolve`, which provides a wrapper (interface) for the `lp_solve` open-source solver. This package's performance is considerably faster, because `lp_solve` solver is implemented in C programming language. However, the same issue as for `linprog` is presented, all objects must be stored at R environment, and the program crashes when trying to allocate large constraints matrix A. The same problem applies to the package `Rglpk`, which uses the GLPK solver.

Further considerations lead us to the conclusion that the most convenient solution would be to store the problem formulation in a separate file, and thus use the full space of the hard drive for storage. Again, sticking only to free solvers, the first suitable candidate is algebraic high-level language MathProg (sometimes referred to as GMPL, GNU Mathematical Programming Language), which is very similar to its proprietary competitor AMPL. The syntax of the language seems to be very attractive from an implementation point of view. Unfortunately, MathProg is very demanding in terms of parsing time and according to the creator of GLPK library Andrew Makhorin: "The MathProg translator is not intended (by design) to process very large models. In such cases, it is better to use the GLPK API or to generate model data, for example in the GLPK LP format, with a specialized program or script". We decide in favor of the CPLEX LP format (`.lp` files, similar to GLPK LP format), because using the GLPK API is meaningful only if the whole routine is adapted to (rewritten in) C language. The chosen approach, however, is not optimal, as long as writing and reading files are a main bottleneck of the routine, and files sizes grow exponentially. The second point is the `.lp` files should be kept in storage, instead of directly be set by C functions of GLPK API. It is also worth to mention the package `glpkAPI` which proposes a "a low level interface to the C API of GLPK". Thus, while extending this paper one can use either native C API of GLPK or mentioned-above package. By default, GLPK uses the primal simplex method. Other available methods include the dual simplex method (both LP and MIP) and the primal/dual interior point method for LP problems.

Finally, the application is organized as follows: The entire source is controlled by Shell script, which consequently calls R scripts and GLPK as a command-line solver. The R code consists of scripts, which generate a scenario tree first as a matrix, and then creates the optimization problem file (in CPLEX LP format). Then the `glpsol` command reads this file and produces the output files with the solution. The size of the intermediate `.lp`, as well as the time of execution of the R scripts and `glpsol` command are tracked. `.lp` files can take gigabytes of storage, thus after successful optimization they are dynamically deleted.

On Figure 5.25, the size of the scenario tree object in R, the size of the `.lp` file and the size of the memory used by the GLPK solver as functions of the number of intervals per variable k are displayed. The thick line indicates the size of available storage, which is the limiting factor. All plots use logarithmic scale. It is clear from these plots, that the memory used increases exponentially, especially for larger planning horizons T. Further, on Figure 5.26 we depict the size of the `.lp` file as a function of the planning horizon T (again, the thick line indicates the storage limit and the plot uses logarithmic scale). It is clear that with the given approach and available equipment, we cannot go further than $T = 8$. For this figure, we only use problems with $k = 2$ number of intervals per variable.

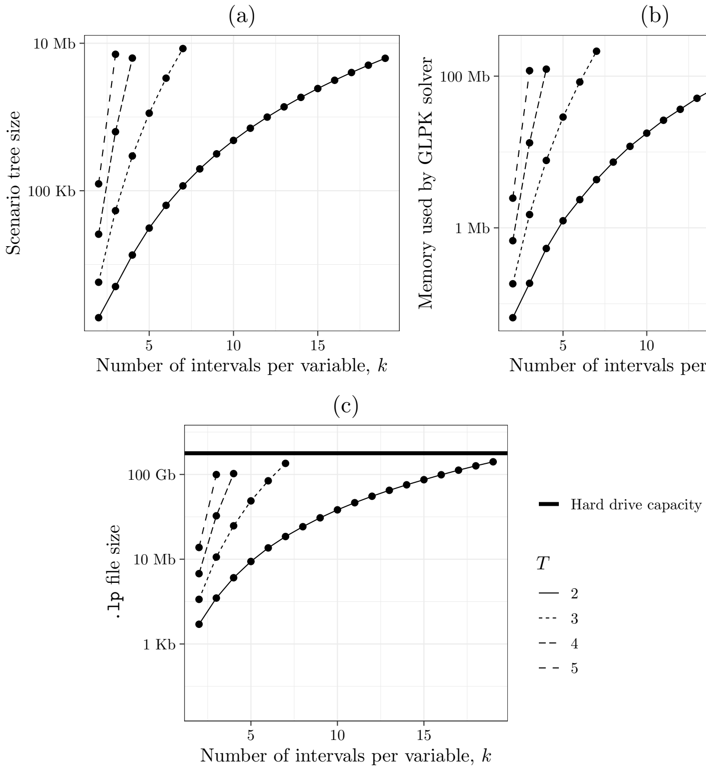

Figure 5.25: Size of the scenario tree object in R (a), memory used by the GLPK solver (b) and size of the `.lp` file (c) depending on the number of intervals per variable k for different planning horizons T.

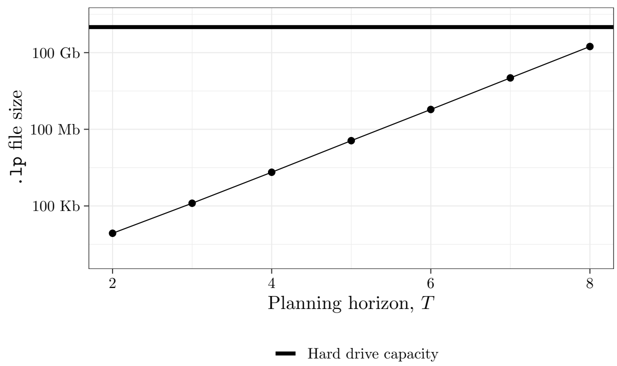

Figure 5.26: Size of the `.lp` file depending on the planning horizon T given $k = 2$.

Figure 5.27 depicts several characteristics related to the time of execution. For subfigures (b) and (c), we again use a logarithmic scale. It is interesting to see, that for odd numbers of intervals per variable the time elapsed on generating scenario trees tends to be greater than for even numbers in (see graph a). This stems from the approximated discrete equivalent in case of odd numbers where an estimation of the vector around zero (central point) is required, which is more significant when compared to the estimation of the other (corner) points. While the time spent for generating scenario trees is important when compared to the time of creating the `.lp` file for the smaller time horizons, we also see peaks at planning horizons $T = 3$ and $T = 5$ in graph (b). The linearity of both functions in (b) and (c) indicates exponential growth of time as the length of the planning horizon and the number of intervals per variable increase.

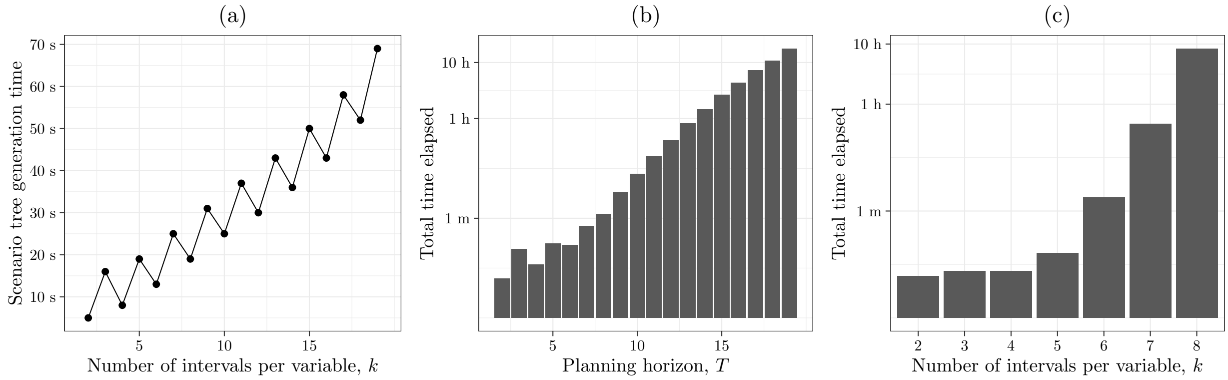

Figure 5.27: Time elapsed for the generation of the scenario tree (a) and total time elapsed for the whole routine (b, c) depending on the number of intervals per variable k (for $T = 2$) and the planning horizon T (for $k = 2$).