University of Groningen

Incentives in the insurance industry

Gorter, Jan Kornelis

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Gorter, J. K. (2013). *Incentives in the insurance industry*. University of Groningen, SOM research school.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 03-06-2022

Incentives in the Insurance Industry

Jan Kornelis Gorter

Publisher: University of Groningen, Groningen, the Netherlands
Printer: Ipskamp Drukkers B.V.
ISBN: 978-90-367-6468-1
eISBN: 978-90-367-6467-4

© 2013 J.K. Gorter

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, now known or hereafter invented, including photocopying or recording, without prior written permission of the author. The views expressed in this thesis cannot be attributed to any of the institutions to which the author is affiliated.

RIJKSUNIVERSITEIT GRONINGEN

Incentives in the Insurance Industry

Proefschrift

ter verkrijging van het doctoraat in de
Economie en Bedrijfskunde
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
donderdag 24 oktober 2013
om 16:15 uur

door

Jan Kornelis Gorter

geboren op 11 mei 1982
te Leiden

Promotores:	Prof. dr. K.H.W. Knot
	Prof. dr. J.A. Bikker
Beoordelingscommissie:	Prof. dr. J.D. Cummins
	Prof. dr. J. de Haan
	Prof. dr. R.J.A. Laeven
	Prof. dr. L. Spierdijk

ACKNOWLEDGEMENTS

Many people have contributed to this thesis. Some have done so directly, others indirectly; numerous have contributed without even knowing it. I would like to express my thanks to all of you, and to some in particular.

When I started my career at De Nederlandsche Bank (DNB) in 2006, DNB had recently merged with the former Pensions and Insurance Chamber and wanted to increase her knowledge of the (Dutch) insurance industry. Klaas Knot and Aerdt Houben gave me the opportunity to contribute to this worthy objective and simultaneously write my thesis. I owe both of them a great deal of gratitude. I am also much indebted to Paul Hilbers and Paul Cavelaars, who allowed me to finish this thesis during turbulent times. While short-term deadlines were piling up, they gave me the opportunity to continue to combine policy work with research.

My supervisors Klaas Knot and Jaap Bikker deserve much appreciation. They gave me full freedom in choosing my topics, and put great trust in me throughout the project, even though not all (self-imposed) deadlines were met. They encouraged me to submit research at an early stage, both to conferences and journals. Jaap was closely involved in the project, both as co-author and as mentor. Despite a busy agenda, Jaap was always eager to discuss research, typically on short notice. It surely helped that our offices were only a few meters apart. Klaas' agenda was even more demanding, His exponential career path brought him a corresponding increase in responsibilities. Nonetheless, he managed to find time to continue to supervise my thesis, which is remarkable. I have fond memories of the discussions Klaas, Jaap and I had, in Amsterdam and The Hague, and in increasingly large offices.

I would also like to use this occasion to show gratitude to Jakob de Haan and Jan Jacobs, both of whom have had a big influence on me. My very first economic research project was with Jakob and Jan. It started as part of an introductory course in econometrics, taught by Jan. The subject was the European Central Bank's monetary policy, a topical issue, even more so these days. Thanks to a joint effort and literally years of patience, we managed to upgrade this research and convert it into a nice publication. This gave me the confidence to pursue a doctorate in economics, though on an entirely different subject. Jakob and Jan also encouraged me to do an internship at DNB, which I can recommend to any aspiring student.

I am grateful to David Cummins, Jakob de Haan, Roger Laeven, and Laura Spierdijk for finding the time to be on my committee and reading my manuscript.

The colleagues at DNB are the best colleagues one could wish for. A special word of thanks goes to everyone from the Strategy department, and also to all the colleagues from Insurance Supervision with whom I have worked closely over the years. Former colleagues who keep in touch are a special breed. Paul, thank you for your friendship and co-authorship.

A warm thank-you goes out to Douwe, Marianna, Marco, Marieke, Rob, Sophie, Reina, Guido, Anje, Marieke, Arina, Pim, Ronald, Marits, Brian, Patricia, Martijn and Mira. We have had some great dinners, parties and holidays together. May there be many more to come!

My ‘clubgenoten’ from JC Amnesia and ‘dispuutsgenoten’ from GHD Babylon deserve special appreciation. Both bonds are from my study years in Groningen and mean a lot to me. Jeroen and Frank, thank you for being my paranymphs.

Family is everything. The sudden loss of my sister Joanne only two and a half months ago came as a serious shock to me and my family, and remains incomprehensible to this day. Dear Joanne, we will remember for the rest of our lifetimes. I truly believe that your spirit lives on, and is with us, also on this day. Corry, Kees, Monica, Ronald, Petra, Frans, and little Marie, thank you for your love and support. Anneke, Arie and Evelien, though I am not married to your wonderful daughter/sister yet, you are already like family to me.

The best is saved for last. Dear Gon, we met here in Groningen ten years ago. I immediately recognized how special you are. Both you and my intuition have not disappointed me ever since. On the contrary, you are the love of my life. In some ways we are different but in others we are remarkably alike. Our relationship is the stable core that we can build on and enjoy life. Let’s continue do that!

CONTENTS

Chapter 1

INTRODUCTION

Over the past years we have seen the worst economic crisis since the 1930s unfold. What started as a seemingly trivial event in the US mortgage market in 2007, i.e. rising delinquencies on subprime mortgages, later developed into a full-fledged global financial crisis with severe economic consequences. In comparison to other types of financial institutions, insurance companies have fared relatively well in the global financial crisis. Indeed, while the majority of banks have received State support - either explicit or implicit - and many pension funds have become underfunded, insurance failures have been much more rare.[1] From this, one may be tempted to infer that insurers are incentivized to stay out of harm's way and act in policyholders' interests. Many policyholders in the Netherlands would not agree, though. In fact, consumer trust in Dutch insurers is quite low these days. This lack of trust is to a large extent the result of the so-called 'woekerpolisaffaire' or 'profiteering policy affair'.

Around 2006, public awareness in the Netherlands increased that various types of unit-linked insurance policies (policies where the value of the policy is linked to investments) carried high, hidden charges. While these unit-linked policies were considered exotic in the early 1990s, at the turn of the century millions of unit-linked policies had been sold, mainly through intermediaries. This impressive rise of the unit-linked market coincided with the equity boom of the late 1990s, which made these policies look cheap. Fiscal incentives also played an important role, as life insurance products were fiscally stimulated compared to straightforward savings and investment products. Insurers and intermediaries had an incentive to sell unit-linked policies, as it was highly profitable to do so (Boot 1995), at least in the medium term. Today, however, the unit-linked portfolios are a millstone around the neck for Dutch life insurers. Billions of euros in cost compensations have already been paid (Verbond van Verzekeraars 2010) and potential mis-selling claims continue to pile up.

The Netherlands is certainly not the only country with a (potential) mis-selling issue in the insurance industry. Another recent case is the United Kingdom (UK), where there continue to be complaints about payment protection insurance (PPI). PPI covers debt repayments if the borrower is unable to meet them in pre-specified situations. According to FSA (2013), PPI was mis-sold on a large scale and lenders have paid £8 billion in redress already. In response to the PPI scandal, the FSA has documented several incentive schemes that increase the risk of mis-selling.

---

[1] Some insurers did experience serious difficulties during the global financial crisis, most notably the American International Group (AIG). The majority of these insurers were brought into trouble not by their insurance business but by their (quasi-)banking activities. According to Geneva Association (2010), over 90% of State support to insurers went to those with significant, failing non-insurance businesses. In the Netherlands, two of the largest insurers are currently state-owned (ASR and Reaal), which is not because these insurers failed but because the banks attached to these insurers were deeply troubled (Fortis and SNS, respectively).

The foregoing suggests, first, that incentives play a dominant role in the insurance industry. This is no surprise: while the global financial crisis has been a watershed event for economic theory, the principle that people respond to incentives stands firm (Stiglitz 2012). Second, it underlines that the ways in which incentives drive economic behavior are not easily predictable. Whereas the favorable crisis track record of insurance companies suggests well-aligned incentives, the profiteering policy affair in the Netherlands and the PPI affair in the UK suggest differently. These affairs point to short-term sales incentives that are not only damaging to policyholders but also pose a threat to the long-term viability of insurance companies.

It is against this backdrop that this thesis analyzes the role and impact of incentives in the (Dutch) insurance industry. The thesis is written from a general perspective, yet all chapters are either inspired by or use data on the Dutch insurance industry, one of the oldest and largest insurance markets in the world. By studying the Dutch insurance industry this thesis aims to enhance our understanding of insurance markets in general.

The remainder of this introductory chapter is structured as follows. To set the stage, Section 1.1 provides a bird's eye view of the Dutch insurance industry. Recent developments in the Dutch insurance industry raise a number of interesting research questions that are discussed in Section 1.2. Section 1.3 sketches the outline of the rest of the thesis.

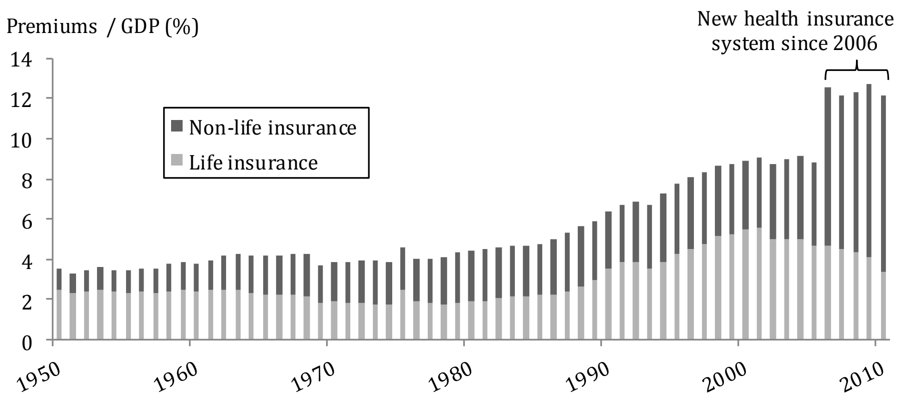
1.1 A BIRD'S EYE VIEW OF THE DUTCH INSURANCE INDUSTRY

Insurance is an important aspect of modern economies. In fact, the first United Nations Conference on Trade and Development (UNCTAD) acknowledged that "[…] a sound national insurance and reinsurance market is an essential characteristic of economic growth" (UNCTAD 1964, p. 55). The Netherlands is a case in point. In the sixteenth and seventeenth century, the Netherlands was the leading economic power in the world (Dutch Golden Age). Much of the Dutch wealth came from overseas trade that was facilitated by marine insurance. The first marine insurance policies were written at the bourse of Antwerp, in the 1530s. A few decades later, fire insurance was introduced, based on mutuality between guild members. The first Dutch non-life insurance company was founded in 1720 (Stad Rotterdam, nowadays part of ASR).

While the history of life insurance in the Netherlands goes all the way back to the Middle Ages, modern life insurance also originated in the Golden Age. In 1671, Dutch statesman and mathematician Johan de Witt published on the valuation and pricing of life annuities. At the time, life annuities were issued by municipalities, in particular within the wealthy province of Holland, to finance public expenditures. Since these annuities were priced age-independently, parents had an incentive to close life insurance not on their own life but on one of their children's lives (Alter and Riley 1986). De Witt understood this incentive and he therefore proposed to link annuity benefits explicitly to the annuitant's remaining life expectancy. De Witt's work is generally seen as the starting point of modern life insurance mathematics (Oosenbrug 2007). The first Dutch life insurance firm was the Hollandse Sociëteit (currently part of Delta Lloyd), founded in 1807.

Since the early start of commercial insurance activity in the Netherlands, almost five hundred years ago, the insurance industry has greatly evolved. Figure 1.1 plots total insurance premiums as a percentage of gross domestic product (GDP) for the relatively recent time period 1950-2010. For decades, the insurance industry developed more or less in pace with overall economic activity in the Netherlands. This changed in the 1980s, wherein total insurance premiums, and especially life insurance premiums, grew faster than GDP. The boom of the life insurance industry continued through the 1990s. Growing prosperity, generous fiscal incentives and high investment returns drove especially the sales of individual life insurance policies to continuously new records. The group life insurance segment (pension contracts) grew more gradually.

Figure 1.1. Total insurance premiums relative to GDP, Netherlands, 1950-2010



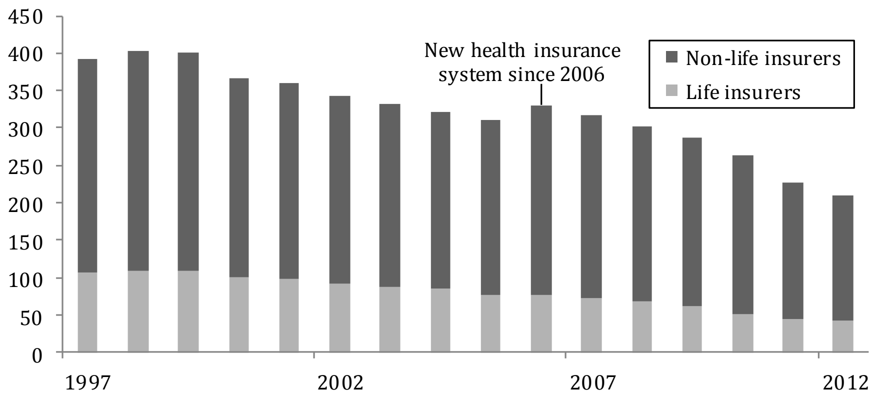
Notes: Data are from Statistics Netherlands and DNB. The insurance premiums are net of reinsurance premiums ceded.

The turning point for the Dutch life insurance industry came around the turn of the century. After years of growth and profitability, total life insurance premiums started to fall, first relative to GDP, and more recently even in absolute terms. In 2001/2002, the dotcom equity bubble burst, which made investment-based life insurance products seem much less attractive than before. At the same time, with the major tax revision of 2001, the tax benefits for new life insurance policies were greatly reduced. The more recent and sharper fall in new life insurance sales was preceded by the Bank Saving Act of 2008 that created a fiscal level playing field for life insurers and banks. As a result of this act, bank and life insurance products have equal fiscal privileges nowadays.

The non-life insurance industry has generally been more stable in terms of premium income. However, with the introduction of the new health insurance system in 2006, non-life insurance premiums more than doubled. At present, the Dutch insurance industry earns annually about €70 billion in premiums (net of reinsurance premiums ceded), which makes it the tenth largest market in the world. Measured per capita, only the Swiss spend more on insurance policies (Swiss Re 2012). Hence, both in absolute and in relative terms the Dutch industry is quite substantial.

While the Dutch insurance industry has grown in size, the number of insurers in the Netherlands has actually decreased (Figure 1.2). Even with the inflow of former health insurance funds in 2006, the number of non-life insurers decreased by almost 40% since 1997. In the life insurance industry, the number of insurers more than halved. This consolidation of the Dutch insurance industry has led to fewer and larger insurance firms.

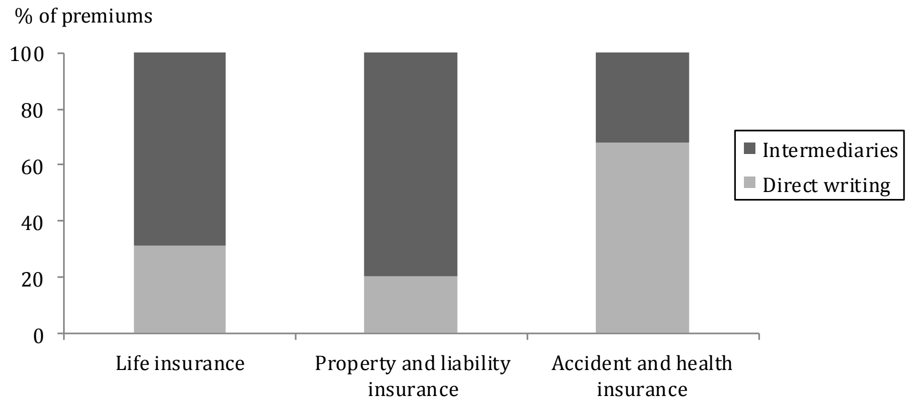
Figure 1.2. Number of Dutch-licensed insurers, 1997-2012



Notes: Data are from DNB. Funeral-in-kind insurers, small mutual non-life insurers exempted from supervision, reinsurers and foreign-licensed insurers are not shown.

Insurance policies are sold either directly, or via intermediaries, such as brokers, agents and banks. Figure 1.3 shows the importance of direct and intermediated sales for different segments of the Dutch insurance industry. In life insurance and also in property and liability insurance (i.e. non-life insurance excluding the line of business accident and health), intermediaries are the main distribution channel. While many large insurance firms have had exclusive contracts with intermediaries for decades, firms are increasingly pursuing so-called multi-channel distribution strategies, with both direct and intermediated sales. Technological developments, such as the growing use of the internet, (tablet) computers and (smart) phones, have facilitated insurers to open up new distribution channels. In accident and health insurance, direct writing is the dominant distribution channel, accounting for 70% of premium income.

Figure 1.3. Distribution-channel shares Dutch insurance industry, 2011



Notes: Data are from Verbond van Verzekeraars (2012). For expositional purposes, non-life insurance is broken down into two parts: accident and health insurance, and property and liability insurance.

1.2 RESEARCH QUESTIONS

A distinctive feature of the Dutch insurance industry is the importance of intermediaries as a distribution channel for insurance products. Intermediaries help customers find the right insurance policy, or bundle of policies, by giving them expert advice. Also outside the insurance industry, i.e. in banking and asset management, retail customers tend to rely on expert advice when making financial decisions. Apparently customers find it difficult and tiresome to evaluate financial products.

Internationally, intermediaries are typically not remunerated directly by their customers but indirectly by financial institutions, through sales commissions (Inderst and Ottaviani 2012a). There is growing concern among policymakers that indirect compensation in the form of commissions leads to biases in financial advice. Already in 2009, the European Commission (2009) expressed doubts over the quality of financial advice, and in particular over the potential conflicts of interests that arise from commission-based remuneration. As financial institutions can influence intermediary advice through commissions, several countries have decided to ban such inducements, at least for certain products. In the Netherlands, commissions on mortgages, life insurance, funeral-in-kind insurance and disability insurance are prohibited per 1 January 2013. Australia and the UK have also decided to ban kickbacks, though on a more limited basis.

Commission bans alter the monetary incentives of intermediaries and are intended to improve the quality of advice. When commissions are banned, intermediaries are directly remunerated by their customers and have an incentive to advice the most suitable product. However, while there is evidence that commission payments bias advice (Anagol, Cole, and Sarkar 2012), there is no evidence that a ban on commissions improves the quality of advice. Understandably, as the commission bans in the Netherlands and other countries are recent, empirical data on the effect of commission bans is at this early stage unavailable. Yet there are already theoretical reasons to fear that commission bans won't improve the quality of financial advice. The crux is that financial institutions also provide advice to customers directly, enabling them to bypass fee-remunerated intermediaries. Whether that is actually a profitable strategy for insurers depends on the incentives. The foregoing raises the following two research questions:

RQ1: *To what extent does a ban on commissions incentivize insurers to bypass intermediaries and provide financial advice directly?*

and

RQ2: *With multiple distribution channels, what is the expected impact of commission bans on the quality of financial advice?*

As described, the insurance industry in the Netherlands is quite sizeable, not only relative to the overall economy but also in absolute terms. Fiscal incentives for life insurance policies and the privatization of former health insurance funds have certainly contributed to this. The large size of the insurance industry is often also attributed to a generally high degree of risk aversion among the Dutch people. Though great aversion to risk could indeed explain a large insurance market, there is little or no direct empirical evidence to substantiate this hypothesis. A key reason for this is that it's typically quite hard to disentangle the impact of risk type and risk preferences on insurance demand. Researchers need detailed information on the pricing practices of insurance companies to be able to come to meaningful conclusions (Chiappori and Salanié 2008). Indeed, risky individuals may be expected to have an incentive to buy more insurance coverage, yet when insurance companies price their policies completely risk-based, this incentive disappears.

Interestingly, in the context of basic health insurance in the Netherlands, insurers are prohibited to price differentiate between low- and high-risk policyholders. This pricing restriction makes the Dutch basic health insurance market close-to-ideal to study insurance demand. While the basic health insurance policy is mandatory for all Dutch residents, it comes with five voluntary deductible alternatives, ranging from €0 till €500 (roughly $650) in stages of €100.

In insurance, the deductible is the amount payable by a policyholder when the insured event occurs. Deductibles alleviate moral hazard that arises from insurance coverage. For research purposes, the policy feature of a voluntary deductible is valuable, as deductible choice data provide insight into people's insurance behavior in the domain of small risks. Policyholders that opt for a high deductible have less insurance coverage, yet with the benefit of a lower insurance premium. Remarkably, ever since the introduction of the new health insurance system in 2006, the large majority of Dutch adults (95% in 2008) refrain from choosing a nonzero deductible. With that, they pass over premium rebates in the order of €200 a year (for the maximum deductible). It is unclear whether this low appetite for higher voluntary deductibles is the result of a high degree of risk aversion, too low premium rebates, or a combination of both. This brings us to the following research question:

RQ3: *How important is risk aversion in explaining the limited appetite for nonzero voluntary deductibles in Dutch basic health insurance?*

The consolidation of the Dutch insurance industry has already been discussed, yet what has been driving this trend towards fewer and larger insurers? An obvious potential reason for consolidation are economies of scale. When scale economies are present, firms have an incentive to engage in mergers and acquisitions, resulting in industry consolidation. Scale economies are present if the unit costs of production decline with firm size, and derive from the spreading of fixed costs over a broader output base. Another potential motivation for consolidation stems from differences in the X-efficiency of insurance firms. X-efficiency reflects managerial ability to drive down production costs, controlled for output volumes and input price levels (Leibenstein 1966). An X-inefficient insurer is a natural takeover target for insurers with more capable management.

Bikker and Van Leuvensteijn (2008) investigate the Dutch life insurance industry and report scale economies and X-inefficiencies. Their results suggest efficiency considerations played a role in the consolidation of the life insurance industry. It is unclear how important efficiency considerations have been in the consolidation of the non-life insurance industry, as this remains to be investigated. Therefore, the fourth research question is:

RQ4: *To what extent was the restructuring of the Dutch non-life sector driven by efficiency considerations?*

This introduction started with the stylized fact that insurers have generally withstood the global financial crisis much better than banks and pension funds. This outcome is often attributed to insurers' more cautious investment strategies. Since (life) insurers and pension funds offer competing products, the question arises why insurers invest more cautiously than pension funds.

According to theory, investors face both risk-management and risk-shifting incentives. Risk-management incentives stem from bankruptcy costs and opportunity costs of missing profitable future investment projects (Mayers and Smith 1987, Smith and Stulz 1985). Risk-shifting incentives originate from an asymmetry in the distribution of profits and losses (Esty 1997). Such an asymmetry occurs clearly in stock-owned firms, as shareholders' upside potential is virtually unlimited, whereas losses are capped by limited liability. But also in pension funds there may be an incentive for risk-shifting, for example between the employer and the employees. The employer may have a preference for more investment risk in the pension scheme so as to lower its annual pension contributions. Accordingly, the fifth and last research question is:

RQ5: *How do incentives affect the investment strategies of insurance firms and pension funds respectively?*

1.3 GOAL AND STRUCTURE

The goal of the next chapters is to examine the five research questions highlighted above. The answers to these questions will give more insight into how incentives have shaped and may continue to shape the Dutch insurance industry. While the unifying theme of this thesis is incentives in insurance markets, the chapters are quite diverse. Chapter 2 investigates how financial advice may develop when commissions are banned, using an extended version of the matching model recently developed by Inderst and Ottaviani (2012b). The key adjustment to the existing model is that both direct and intermediary advice are allowed for, instead of only intermediary advice.

Chapters 3-5 are empirical in nature. Chapter 3 investigates the importance of risk preferences in explaining deductible choice in Dutch basic health insurance. The combination of population data and DNB Household Survey (DHS) data makes the analysis unique. Chapter 4 investigates the incentives behind the restructuring of the Dutch non-life insurance industry over the period 1995-2005. This period of restructuring was preceded by a series of European deregulations, which suggests efficiency considerations played a pivotal role. Chapter 5 explores the investment risk taking of Dutch insurers and defined benefit (DB) pension funds over the volatile period 1995-2009. In 2009, these institutional investors collectively managed over €1,000 billion in assets, far exceeding Dutch GDP in that year (less than €600 billion). The incentives of these investors are therefore important, not only for individual policyholders and pension plan participants, but also from a macro-prudential policy perspective. Chapter 6 returns to the research questions formulated in this introductory chapter and discusses the implications of the research. Chapter 7 provides a Dutch summary of the thesis.

Chapter 2

ADVICE INCENTIVES WHEN COMMISSIONS ARE BANNED[1]

[1] This chapter is based on Gorter (2012).

[T]he history of retail financial services over the last 20 years has not been a happy one: punctuated with too many waves of mis-selling – large-scale customer detriment followed by large imposed compensation.

Adair Turner, 2011

2.1 INTRODUCTION

Sound financial advice is key to the well-functioning of financial markets. Financial decisions tend to be complex and errors can have serious consequences. Retail customers find it particularly hard to evaluate financial products. They therefore tend to rely on expert advice when making financial decisions.[1] Unfortunately, as the opening quote of Adair Turner highlights, expert advice has not prevented mis-selling in retail financial services.

In light of this, several countries have recently adopted laws to improve the quality of financial advice. Notably, as from 2013, independent financial advisers in Australia, the Netherlands and the UK are prohibited to accept commissions from financial institutions.[2] At the European level, the European Commission (EC) has proposed to ban inducements in the context of the review of the Markets in Financial Instruments Directive (EC 2011) and the Insurance Mediation Directive (EC 2012). With a ban on commissions, financial advisers are remunerated directly by their customers, through an hourly or fixed fee. It is envisaged that fee-based remuneration will lead to more suitable advice, as there is no incentive to advise a particular, high commission product.

Against this background, this chapter investigates the incentives behind advice when commission payments are banned. To this end, the Inderst and Ottaviani (2012b) framework is extended by also allowing for direct advice by financial institutions. In an innovative article, Inderst and Ottaviani show why specialized financial advisers (henceforth intermediaries) are commonly remunerated through commissions rather than through fees. They show that commission-based remuneration enables financial institutions to benefit more from so-called naïve customers who do not properly take into account the incentives behind advice. As commissions undermine the quality of financial advice and therefore total welfare, the authors

[1] According to a survey by Hung, Clancy, Dominitz, Talley, Berrebi, and Suvankolov (2008), over seventy percent of US retail investors consult an adviser before buying shares. Chater, Huck, and Inderst (2010) survey recent buyers of financial products across eight EU countries and find that almost sixty percent were strongly influenced by an adviser before buying shares.

[2] While these countries have all decided to ban commissions on retail financial products, there are differences. In Australia, the ban holds for superannuation and investment products, and not for pure protection products and mortgages. Moreover, the ban commenced on 1 July 2012, yet compliance will be mandatory only from 1 July 2013 (Commonwealth Treasury 2012). In the Netherlands, the commission ban holds for complex financial products, such as mortgages, life insurance, funeral insurance and disability insurance. Other insurance products are outside the scope of the commission ban (Minister van Financiën 2011). In the UK, the commission ban holds only for advised sales, not for non-advised sales. Protection-only insurance products, e.g. mortality insurance and property and casualty insurance, and mortgages are currently outside the scope of the commission ban (FSA 2009).

conclude there is an economic rationale for commission bans. We establish that when financial institutions have the opportunity to advise customers directly, which is typically the case in practice, there is a genuine risk that commission bans won't improve the quality of financial advice.

In our model there are two types of advisers: a financial institution and an intermediary. While advice is non-verifiable, advisers incur costs in case of unsuitable advice. These costs may be interpreted as reputational costs, yet we assume, for concreteness, that these represent regulatory fines. In terms of remuneration, we specify that intermediaries are remunerated through fees, as our analysis focuses on financial advice when commissions are prohibited. Besides conditional fixed transfers between the financial institution and the intermediary are also prohibited. The financial institution and the intermediary operate at arm's length, which is different from Inderst and Ottaviani's (2012b) baseline model, where these parties cooperate and maximize joint profits. The model treats financial products as experience goods, since customers sooner or later find out whether they have bought the right product or not. Ex ante, however, customers are unaware which product is most suitable. Moreover, customers differ in their understanding of incentives that advisers have. While alert customers understand that the financial institution has an incentive to recommend the most profitable product, naïve customers believe that profitability does not play a role and that financial advisers always advise the most suitable product.

By way of preview, our main findings are as follows. In equilibrium, financial institutions find it optimal to sell to alert customers via intermediaries, and to naïve customers directly. As profit maximization implies that direct advice is tilted to the most profitable products, naïve customers unfortunately receive distorted advice. The root cause of this inefficient market outcome is that naïve customers mistakenly believe that the quality of direct advice is equivalent to that of intermediary advice. This misunderstanding of advice incentives makes it more profitable to financial institutions to deal with naïve customers directly. Alert customers obtain sound advice, however, as their willingness to pay for financial products and advice depends directly on the quality of financial advice. The model's equilibrium is incentive compatible: neither alert nor naïve customers have an incentive to switch from direct to intermediary advice, or vice versa.

In terms of advice quality, our equilibrium is identical to Inderst and Ottaviani's (2012b) baseline model equilibrium *with* commission payments and the possibility of indirect price discrimination (see Proposition 3 in their article). This is an important result, as it indicates that commission bans may not improve the quality of financial advice after all. The intuition behind the similarity of equilibriums is relatively clear when it comes to naïve customers. In our model these customers receive direct advice, which is equivalent to intermediary advice that is steered by the financial institution through commissions, as analyzed by Inderst and Ottaviani. With respect to alert customers, the correspondence is more subtle. In Inderst and Ottaviani's equilibrium, alert customers are convinced that advice is sound as they are only charged a fee for advice. By contrast, in our equilibrium, alert customers are charged both a fee for advice and a positive premium product price. Since commission payments are forbidden, alert customers rightly do not associate a positive product price with substandard advice quality.

The model's equilibrium thus depends crucially on customer rationality. An important question is then to what extent customers fail to take the incentives behind advice into account. In a large-scale survey among recent purchasers of retail investment products in Europe, Chater et al. (2010) find that the majority of respondents are unaware of potential conflicts of interest. Of the respondents who were advised directly by product provider staff, more than half perceived the advice given to be completely independent and unbiased. In a study of stock recommendations, Malmendier and Shantikumar (2007) find that small investors follow analyst recommendations literally, even though many analysts are affiliated to underwriters and thus not independent. The authors conclude that naiveté about advice distortions is a realistic explanation of this great willingness to follow advice. Similarly, Cain, Loewenstein, and Moore (2005) report experimental evidence that many participants have an excessive tendency to follow advice, even if distorting incentives behind advice are disclosed. By adopting the distinction between naïve and alert customers, this research fits within a growing stream of industrial organization research that takes consumer biases into account (see Ellison 2006 for a relatively recent review). Related research on consumer biases in financial services is Carlin (2009). In his model, however, customers vary in their understanding of prices, not in terms of their understanding of advice incentives. Moreover this chapter examines horizontally-differentiated products, while Carlin (2009) investigates homogenous goods.

Information provision by product providers is also analyzed by, among others, Lewis and Sappington (1994), Moscarini and Ottaviani (2001), Bolton, Freixas, and Shapiro (2007), and Bar-Isaac, Caruana, and Cuñat (2010). Of these, only Bolton et al. (2007) focus on the financial services industry. Their key finding is that competition between financial institutions can lead to full credible information provision, even with only a small reputation cost. This result is comparable to our result with alert customers, whereby in our model reliable information provision is not the result of competition but of the presence of a trustworthy intermediary. Inderst and Ottaviani (2009) also analyze situations where product advice is tied to the sale, but they focus on how financial institutions can optimally compensate internal sales agents. The benefits of steep sales incentives need to be balanced with the costs, that is, expected losses from selling unsuitable products.

In addition to financial advice, this chapter sheds light on distribution strategies in financial services. For example, in insurance, there are two types of distribution strategies. So-called independent-agency insurers sell via intermediaries, whereas so-called direct writers sell directly. While independent-agency insurers have higher costs, both types of insurers have coexisted in insurance markets for decades. The dominant explanation for this coexistence is that independent-agency insurers provide higher-quality services, which explains their higher costs (Berger, Cummins, and Weiss 1997). Our analysis reveals that the choice of distribution channel may also depend on intermediaries' remuneration. When commissions are banned and intermediaries are directly remunerated by their customers, it is attractive to follow a so-called multichannel distribution strategy, that is, selling to naïve customers directly and to alert customers via an intermediary.

The rest of the chapter is organized as follows. Section 2.2 describes the model that allows for both direct and intermediated sales. Section 2.3 investigates what determines advice and how the quality of advice affects profits, consumer surplus and total welfare. To focus initially on direct sales, Section 2.4 examines market equilibriums without intermediary involvement. Section 2.5 analyzes the full model, whereby the financial institution has two distribution channels to choose from: direct and intermediated sales. Section 2.6 gives policy implications, and Section 2.7 concludes. Appendix 2.A gives the proofs of the propositions. An illustration of the model without intermediary involvement is given in Appendix 2.B.

2.2 MODEL OUTLINE

The model builds on Inderst and Ottaviani's (2012b) baseline model. There are three strategic players: a financial institution, an intermediary and a customer. The main difference with the existing model is that we allow for two sources of financial advice: that is, both financial institutions and intermediaries advise on product suitability.

The financial institution has two financial products on offer, products *A* and *B*. These products are imperfect substitutes. Product *A* is a premium product that is only provided by the modeled financial institution; product *B* is a plain-vanilla product that is provided by many other firms and is competitively priced. For simplicity, the production costs of the two products are normalized to zero. Since *B* is also provided by other firms with equal cost levels, the price of *B* is zero.[3] The financial institution does have price-setting power over product *A*, however. A high price for product *A* raises the profitability differential between *A* and *B*, which is key to the equilibrium outcome. In the insurance domain, product *A* could be a unit-linked life insurance

[3] The results are identical when the financial institution offers only product *A* (it is in this respect a monopolist) and the other firms provide product *B*.

policy with interest rate guarantees included, whereas product B could be a basic fixed annuity. The difference between A and B can also be more subtle. For example, in banking, both products could be fixed rate mortgages, yet A offers more flexibility than B in terms of monthly payments.

While the model is specifically related to financial services, the analysis of direct advice by product providers has wider applicability. Indeed, in any market where the seller has considerable informational advantage over the buyer, the seller does not only sell the product but typically also provides product advice. Examples of such markets outside financial services are consumer electronics, residential real estate and the automotive industry.

2.2.1 Customers

There are N customers in the market. For simplicity, and in line with the related literature, we assume that each customer buys only one product. There are two types of customers, α and β. Customers are unaware of their type, yet know that in the population a fraction q_0 is of type α and $1 - q_0$ is of type β. For concreteness and without loss of generality we assume $q_0 = \frac{1}{2}$.

A key element of the model is match suitability. When type α (β) customers buy product A (B), they are matched with the product that matches their preferences and achieve high utility u_h. On the other hand, when type α (β) customers buy product B (A), we speak of unsuitable choice, and as result customers achieve low utility u_l. Since suitable matches lead to higher utility, we have $u_h > u_l$. For further reference we define $\Delta_u \equiv u_h - u_l$. Since customers are expected utility maximizers, they prefer u_h over u_l. Accordingly, products A and B are horizontally differentiated in the sense of Hotelling (1929). At equal prices, neither A nor B is unanimously preferred by the customers. Without advice, customers make their product choice randomly and have expected utility $u_0 = u_l + q_0\Delta_u = u_l + \frac{1}{2}\Delta_u$.

Besides product preference differences, customers also vary in terms of rationality. There are alert and naïve customers. While the former are aware of possible incentives behind product advice, the latter completely ignore this possibility. Note that no assumption is made about the relationship between customer rationality and customer preferences.

2.2.2 Advisers

Customers can consult the financial institution or the intermediary to obtain information on product suitability. We assume that the quality of information of both advisers is the same. There can be a difference in advice quality, however, namely when advisers face dissimilar advice incentives. We delay a discussion of the particular incentives behind advice to Section 2.3.

When consulted by a customer, an adviser forms a posterior belief q that the respective customer is of type α, which implies a belief $1 - q$ that this customer is of type β. With perfect information, q is either 0 or 1, and an adviser is certain which product is optimal and which is not. With imperfect information, however, q may be any value on the interval [0,1]. Consequently, an imperfectly informed adviser is never fully sure which product is most suitable. The information quality of advisers is reflected by the cumulative distribution of the posterior belief, $G(q)$, which is assumed to be exogenous and known to the customers. The corresponding density function is $g(q) > 0$ for $q \in [0,1]$, which implies imperfect information. Though imperfectly informed, as a rule advisers have an informational advantage over their customers with respect to product suitability (the exception to the rule is when the posterior belief is equal to the prior, i.e. $q = q_0 = \frac{1}{2}$).

Product suitability beliefs are formed in practice through cus**t**omer profiling. Advisers tend to ask their clients about their objectives, income, degree of risk aversion, investment horizon, et cetera. Combining such customer information with product information gives an adviser knowledge about product suitability. For example, when the customer is highly risk-averse and needs a financial product to pay off a mortgage in thirty years' time, simple savings products are more suitable than investment products with a variable rate of return.

The financial institution has three ways to sell its products. It can sell directly without advice, sell directly with advice, and/or sell via the intermediary with advice. The first option leads to a so-called non-advised sale.[4] In the model, non-advised sales lead to zero profits, since customers a priori do not have a preference for *A* over *B*, and *B* is competitively priced at zero marginal costs. This implication of the model echoes Bester's (1998) result that information provision increases the market power of firms, that is to say, the ability to set prices above marginal costs. Hence, non-advised sales are typically not in the interest of the financial institution.

With advised sales, customers receive suitability advice either from the financial institution or the intermediary. In the model, both types of advisers face a direct incentive to provide suitable advice. Specifically, in case of mismatch, an adviser's pay-outs are reduced by $\rho > 0$ (cf. Bolton et al. 2007). This parameter ρ may be interpreted as a loss of business from word-of-mouth that the adviser gives biased advice. ρ may also be interpreted as representing regulatory fines for providing unsuitable advice to customers. For concreteness, we use the latter interpretation of ρ, i.e. a regulatory fine in case of unsuitable choice. Note that an adviser is always fined for a mismatch, even if it is unintentional, that is, caused by imperfect information.

The financial institution's aim is to maximize its total profits. As discussed, profitability must come from advised sales. Denote the profits from direct and intermediated advised sales

[4] FSA (2006a) explains the distinction between advised and non-advised sales in insurance.

π^D and π^I, respectively. Total profits are then $\pi = \pi^D + \pi^I$. While product B's price is fixed at marginal costs, the financial institution has price-setting power over advanced product A. Denote A's direct advised sales price P^D, and the intermediary price P^I. In the baseline model, the product provider provides a take-it-or-leave-it offer to the intermediary. As a robustness exercise, we verify our results when it is the other way around, that is, when the intermediary provides a take-it-or-leave-it offer to the financial institution.

The intermediary is remunerated through a fee for advice $f \geq 0$. This fee is directly paid by the customer. Indeed, other than in Inderst and Ottaviani (2012b), indirect compensation of the intermediary through commissions is prohibited in our model. This allows us to focus on market outcomes when commission bans are effective. The intermediary aims to maximize its profits F. Like the financial institution, the intermediary has zero costs. Since the intermediary is directly remunerated, the quality of advice and the size of the regulatory fines do not depend on f. As a result, the intermediary has an incentive to set f as high as possible.

2.2.3 Timeline

The timing of the game is as follows. In stage $t = 1$, the financial institution sets product A's advised-sales prices, i.e., P^D and P^I. The non-advised sales price of product A is always zero. Based on the intermediated sales price P^I, in $t = 2$, the intermediary chooses its advisory fee f. In $t = 3$, the customer decides between non-advised sales, direct advised sales and intermediated sales. For simplicity, we do not allow for switching between these sales channels after $t = 3$.[5]

When the customer opts for no advice, all payoffs are immediately realized. In this case the financial institution and the adviser earn zero profits and the customers receive payoff u_0. In case of advised sales, however, the game proceeds. When the customer chooses for direct advice from the financial institution, in $t = 4$ the financial institution privately receives information on the customer's type, represented in the model by posterior belief q. Using this information, in $t = 5$ the financial institution recommends the customer which product to buy, A or B. In $t = 6$, the customer chooses and, subsequently, all payoffs are realized. In $t = 3$, the customer may also choose for intermediary advice, in exchange for fee f. After the fee is paid, in $t = 4$, the adviser obtains information on the customer's type, represented by its posterior belief q. Using this

[5] This is less restrictive as it may seem. Since the financial institution has a monopoly over product A, it can easily prevent switching between direct and intermediary advice. Non-advised sales are potentially more troublesome. Indeed, when switching would be allowed for, customers could obtain financial advice, directly or indirectly, and subsequently buy the advised product via the non-advised sales channel at zero costs. Such switching behavior can also be prevented by the financial institution. Indeed, it could simply cease non-advised sales of premium product A. This would not materially change the model, as non-advised sales would still yield expected utility u_0.

information, the intermediary gives a product recommendation in $t = 5$. In $t = 6$, the customer chooses its product and all payoffs are realized. In line with the related literature, all players in the game are risk neutral and payoffs are not discounted.[6]

2.3 ADVICE, PROFITS AND SURPLUS

Advice. When the financial institution gives advice, it has a suitability concern which is captured by parameter ρ. Choosing between products A and B, the institution advises A if

$$P^D - \rho(1-q) \geq -\rho q.$$

For convenience, define cutoff q^* that determines when A is advised. This is the case when $q \geq q^*$, with

$$q^* = \frac{1}{2} - \frac{P^D}{2\rho}, \quad \text{for } P^D < \rho, \quad \text{and}$$
$$q^* = 0, \quad \text{for } P^D \geq \rho. \tag{2.1}$$

For $P^D < \rho$, (1) implies that the higher the price of A (P^D), the lower q^* and the more likely it is that A is advised. This effect of P^D on product advice is alleviated by regulatory fines ρ, which are payable in case of mismatch. Nonetheless, for $P^D > 0$ advice cutoff $q^* < \frac{1}{2}$, and advice is tilted towards the more profitable premium product. Indeed, when the financial institution assesses, for instance, that A and B are equally suitable ($q = \frac{1}{2}$), it will actually advise product A if $P^D > 0$. When $P^D \geq \rho$, advice becomes completely uninformative, as $q^* = 0$ and product A is always advised.

The intermediary is another source of advice in the model. As a result of its direct remuneration, the intermediary provides sound advice and thus applies advice cutoff $q^* = \frac{1}{2}$.

After customers have received advice in $t = 5$, they may possibly deviate from the advice given in $t = 6$. We will show in Sections 2.4 and 2.5, however, that in equilibrium, customers always follow financial advice. The intuition is that prices P^D and P^I and fee f are chosen in such a way that customers find it optimal to receive *and* follow advice, as profits are made from advised sales only. In the remainder of this section, we derive equations for profits, consumer

[6] The assumption of risk-neutral preferences simplifies the model and focuses attention to financial advice. Given that in practice people tend to be risk averse, it would be interesting to explore the model with risk-averse players. We expect that risk-averse customers have a lower willingness to pay for advice than risk-neutral customers, as product suitability remains uncertain, even after having received advice. A full derivation with risk-averse preferences is beyond the scope of this chapter, however.

surplus and total surplus, all under the premise that product advice is followed. Again, the validity of this premise is verified in the next two sections.

Profits. Recall that profits originate from advised sales. Let ε be the fraction of customers who buy without advice, γ the fraction of customers who opt for intermediary advice, and $1-\gamma-\varepsilon$ the fraction that opt for direct advice. Direct sales profits are then $\pi^D = N(1-\gamma-\varepsilon)\tilde{\pi}^D$, where $\tilde{\pi}^D$ are the respective profits per customer. Profits from intermediated sales are $\pi^I = N\gamma\tilde{\pi}^I$, where $\tilde{\pi}^I$ are the corresponding profits per customer. Taking π^D and π^I together, total profits are

$$\pi = N[(1-\gamma-\varepsilon)\tilde{\pi}^D + \gamma\tilde{\pi}^I]. \tag{2.2}$$

Not all customers who obtain direct advice are profitable to the financial institution. Indeed, of the customers who obtain direct advice, only a fraction $1-G(q^*)$ buys the premium product, at price P^D. Hence, given zero production and advisory costs, we have per customer profits

$$\tilde{\pi}^D = P^D[1-G(q^*)] - \rho UC(q^*), \tag{2.3}$$

where

$$UC(q^*) = \int_0^{q^*} q dG(q) + \int_{q^*}^{1} (1-q) dG(q).$$

$UC(q^*)$ is the ex ante probability of *un*suitable choice, which is minimized for $q^* = \frac{1}{2}$.[7] It depends on the information quality to what extent unsuitable choice can be completely eliminated. In case of mismatch, the financial institution needs to pay ρ to the regulator, which is transferred from the adviser to the general public.

With intermediated sales, a fraction $1-G\left(\frac{1}{2}\right)$ of the customers buys the premium product at price P^I. In this case, regulatory fines are payable by the intermediary. Per customer profits from intermediated sales are therefore

$$\tilde{\pi}^I = P^I\left[1-G\left(\frac{1}{2}\right)\right]. \tag{2.4}$$

---

[7] By Leibniz's rule, we have $UC'(q^*) = (2q^*-1)g(q^*)$. Hence, the only potential minimum of $UC(q^*)$ is at $q^* = \frac{1}{2}$. Since $UC''\left(\frac{1}{2}\right) > 0$, $q^* = \frac{1}{2}$ minimizes unsuitable choice.

The intermediary's profits are $F = N\gamma\tilde{F}$, where net earnings per customer $\tilde{F}$ are

$$\tilde{F} = f - \rho UC\left(\tfrac{1}{2}\right). \tag{2.5}$$

Consumer surplus. When customers buy without advice, their choice is based on the prior that half of the population is best off with product A, and the other half is best off with product B. Hence, without additional information, customers have a fifty percent probability of buying the right product (recall $q_0 = \frac{1}{2}$). Non-advised sales yield consumer surplus $N\varepsilon u_0$, with $u_0 = u_l + \frac{1}{2}\Delta_u$.

Denote consumer surplus from direct advised sales $CS^D = N(1-\gamma-\varepsilon)\widetilde{CS}^D$, where the surplus per customer is $\widetilde{CS}^D$. As customers follow advice, we have

$$\widetilde{CS}^D = u_l + \Delta_u[1 - UC(q^*)] - P^D[1 - G(q^*)]. \tag{2.6}$$

With intermediated sales, consumer surplus is $CS^I = N\gamma\widetilde{CS}^I$, where every customer obtains

$$\widetilde{CS}^I = u_l + \Delta_u\left[1 - UC\left(\tfrac{1}{2}\right)\right] - P^I\left[1 - G\left(\tfrac{1}{2}\right)\right] - f. \tag{2.7}$$

Total surplus. Non-advised sales yield no profits, so the contributed to welfare is consumer surplus $N\varepsilon u_0$. In case of advised sales, total surplus is potentially greater. Let TS^D be total surplus created by direct sales with advice. Then we have $TS^D = N(1-\gamma-\varepsilon)\,\widetilde{TS}^D$, where $\widetilde{TS}^D$ is the surplus per direct customer. Adding (2.3), (2.6) and the regulatory fines, we have

$$\widetilde{TS}^D = \tilde{\pi}^D + \widetilde{CS}^D + \rho UC(q^*) = u_l + \Delta_u[1 - UC(q^*)]. \tag{2.8}$$

Similarly, total surplus created by intermediated sales is $TS^I = N\gamma\,\widetilde{TS}^I$, where

$$\widetilde{TS}^I = \tilde{\pi}^I + \widetilde{CS}^I + \tilde{F} + \rho UC\left(\tfrac{1}{2}\right) = u_l + \Delta_u\left[1 - UC\left(\tfrac{1}{2}\right)\right]. \tag{2.9}$$

Since unsuitable choice is minimized with $q^* = \frac{1}{2}$, (2.9) implies that total welfare is maximized when all sales are done via the intermediary, that is $\gamma = 1$. We will show in Section 2.5 that in equilibrium $\gamma \neq 1$.

2.4 DIRECT ADVICE

As the model is quite extensive, it is useful to concentrate on direct sales first. To this end, we assume that the financial institution interacts only directly with its customers ($\gamma = 0$). This has the following consequences for the timing of the game. In $t = 1$, the financial institution sets just A's direct price P^D, instead of two prices. In $t = 2$ nothing happens and in $t = 3$ the customer chooses between no advice and direct advice as $\gamma = 0$. The equilibrium results developed in this section will be useful in Section 2.5, where we analyze the full model with endogenous γ.

Below, in Section 2.4.1, it is first assumed that all customers are alert about the incentives behind advice. Conversely, in Section 2.4.2, all customers are assumed to be naïve and believe that advice is always unbiased. Both assumptions are extreme. In the last subsection, we allow for a mix of alert and naïve customers, which is particularly relevant when it is impossible to price discriminate between customers.

2.4.1 Alert customers

Alert customers understand that advice is potentially biased when one product is more profitable to the financial institution than the other. While such alertness seem to be at odds with extant empirical evidence on typical consumer behavior in financial services, as partly described in Section 2.1, this subsection helps to sheds light on a world wherein customers would be a relatively strong counterforce to profit-maximizing financial institutions.

Customer participation. Recall that the financial institution advises product A if $q \geq q^*$. In $t = 6$, alert customers follow advice to buy A if the expected payoff is greater than or equal to buying B (at zero costs) while A is advised, that is,

$$u_l + \Delta_u \int_{q^*}^{1} q \frac{dG(q)}{1 - G(q^*)} - P^D \geq u_l + \Delta_u \int_{q^*}^{1} (1 - q) \frac{dG(q)}{1 - G(q^*)}. \tag{2.10}$$

Equation (2.10) can be rewritten to the participation constraint

$$P^D \leq \Delta_u V(q^*), \tag{2.11}$$

where

$$V(q^*) = \int_{q^*}^{1} (2q-1)\frac{dG(q)}{1-G(q^*)}.$$

When $q^* = 0$ and product A is always recommended, advice is uninformative, $V(0) = 0$ and alert customers have zero willingness to pay for product A. Their willingness to pay increases, however, when q^* gets closer to $\frac{1}{2}$, and advice becomes increasingly informative.[8] Note that since q^* and P^D are negatively related for interior cutoff values ($q^* > 0$), by (2.1), the financial institution's pricing power is directly restricted by (2.11). Raising the price of A actually reduces alert customers' willingness to pay for A.

The financial institution advises to buy B in $t = 5$ when $q < q^*$. This advice is followed when the expected value of buying B given the advice to buy B is larger than the expected value of buying A given the advice to buy B:

$$u_l + \Delta_u \int_{0}^{q^*} q\frac{dG(q)}{G(q^*)} - P^D \leq u_l + \Delta_u \int_{0}^{q^*} (1-q)\frac{dG(q)}{G(q^*)},$$

which always holds. Hence, advice to buy B is followed by the customer.

In $t = 3$, when choosing between advised and non-advised sales, an alert customer chooses the former when

$$u_l + \Delta_u \int_{0}^{q^*} (1-q)dG(q) + \int_{q^*}^{1} (\Delta_u q - P^D)\, dG(q) \geq u_0, \tag{2.12}$$

where $u_0 = u_l + q_0\Delta_u$ is the expected payoff without advice. By Bayes' law, the expectation of posterior belief q is equal to prior q_0, or $q_0 = \int_0^1 (q)dG(q)$. This implies $u_0 = u_l + \Delta_u \int_0^{q^*}(q)dG(q) + \Delta_u \int_{q^*}^1 (q)dG(q)$. Substituting for u_0 in (2.12) yields

$$P^D \leq \int_{0}^{q^*} (1-2q)\frac{dG(q)}{1-G(q^*)} = \Delta_u V(q^*). \tag{2.13}$$

[8] This is shown in Appendix 2.A.1. Crucial in this respect is the postulation that $g(q) > 0$ for $q \in [0,1]$. If $g(q) = 0$ for any q in the respective interval, an increase in q^* does not necessarily affect the willingness to pay for product A.

As (2.13) is identical to (2.11), and since advice to buy product B is always followed, we conclude that direct financial advice is always followed. Intuitively, both in $t = 3$ and in $t = 6$, alert customers evaluate the value of product advice on the basis of cutoff q^* and distribution $G(q^*)$. So, if the value of advice is deemed insufficient in $t = 6$, that is if (2.11) is not satisfied, it will also be deemed insufficient in $t = 3$, when (2.13) is not satisfied. Note the actual advice in $t = 5$ does remain informative, though, as it informs customers whether $q < q^*$ or $q \geq q^*$.

Financial institution participation. The financial institution provides advice only if the resulting profits are non-negative. Though production costs are set to zero, profits can still be negative because of regulatory fines. If the expected profits are indeed negative, the institution decides either to sell exclusively without advice ($\varepsilon = 1$) or leave the market altogether; in both cases total profits are zero. Using (2.2) and (2.3), the financial institution's participation constraint is

$$P^D[1 - G(q^*)] \geq \rho UC(q^*). \tag{2.14}$$

Equilibrium. In $t = 1$, the financial institution sets P^D so as to maximize total profits in (2.2), with $\gamma = 0$. It is shown in Appendix A.1 that total profits are increasing in P^D as long as customer participation constraint (2.11) is satisfied. Accordingly, the financial institution will raise A's price until (2.11) becomes binding. Therefore, provided that (2.14) holds, the equilibrium direct sales price of A with alert consumers, $P_A^{D,EQ}$, is

$$P_A^{D,EQ} = \Delta_u V\left(q_A^{*,EQ}\right), \tag{2.15}$$

where $q_A^{*,EQ}$ is the equilibrium advice cutoff with alert customers. Combining (2.15) with (2.1) pins down a unique equilibrium, with $P_A^{D,EQ} > 0$ and cutoff $0 < q_A^{*,EQ} < \frac{1}{2}$.

<u>Proposition 2.1</u>. *On condition that (2.14) holds and it is thus attractive to provide direct advice, there exists a unique equilibrium where the financial institution uses advice cutoff* $q_A^{*,EQ}$, *and charges alert customers* $P_A^{D,EQ} = \Delta_u V\left(q_A^{*,EQ}\right)$. $q_A^{*,EQ}$ *is the solution to* $\rho(1 - 2q_A^{*,EQ}) = \Delta_u V\left(q_A^{*,EQ}\right)$. *Since* $0 < q_A^{*,EQ} < \frac{1}{2}$, *financial advice is tilted towards recommending product A.*

A formal proof of Proposition 2.1 is in Appendix 2.A.1. Note that with alert customers, advice remains informative in equilibrium as $q_A^{*,EQ} > 0$. In fact, the equilibrium price that alert customers pay for A is exactly equal to the added value from advice in expected utility terms. To see this, substituting $P_A^{D,EQ}$ and $q_A^{*,EQ}$ into (2.6), we have

$$\widetilde{CS}^{D,EQ} = u_l + \Delta_u\left[1 - UC\left(q_A^{*,EQ}\right)\right] - \Delta_u \int_{q_A^{*,EQ}}^{1} (2q-1)dG(q) = u_0,$$

where we have used $\int_{q^*}^{1}(2q-1)dG(q) = \frac{1}{2} - UC(q^*)$. Since all customers opt for advice in equilibrium (i.e. $\varepsilon = 0$), total consumer surplus is $CS^{D,EQ} = Nu_0$, which is equal to consumer plus with non-advised sales only (i.e. $\varepsilon = 1$). Hence, alert customers are equally well off with and without information, as the financial institution is able to capture the full welfare increase due to its price-setting power. Total surplus increases with information, thanks to a greater likelihood of suitable choice. Indeed, given that $q_A^{*,EQ} > 0$, total surplus is

$$TS^{D,EQ} = N\widetilde{TS}^{D,EQ} = N\left(u_l + \Delta_u\left[1 - UC\left(q_A^{*,EQ}\right)\right]\right) > Nu_0.$$

The next subsection shows that with naïve customers, a completely different equilibrium arises.

2.4.2 Naïve customers

Here we analyze market outcomes when customers are naïve in the sense that they do not recognize that a higher price for product A, and with that a greater price differential with product B, reduces the quality of advice.

Customer participation. Regardless of price P^D, naïve customers expect sound advice and thus belief that the advice cutoff is $\frac{1}{2}$, which may well deviate from the actual cutoff q^*. In $t = 6$, naïve customers are willing to follow advice to buy product A if

$$P^D \leq \Delta_u \int_{1/2}^{1} (2q-1)\frac{dG(q)}{1 - G(1/2)} = \Delta_u V\left(\tfrac{1}{2}\right). \tag{2.16}$$

Recall that the decision between advised and non-advised sales is made in $t = 3$. Naïve customers make a similar assessment as analyzed before for alert customers, with the only difference that former expect balanced advice regardless. Consequently, using advice cutoff $\frac{1}{2}$ in (2.13), we have that alert customers opt for direct advice if (2.16) is satisfied.

Equilibrium. Provided that firm participation constraint (2.14) holds, profits are increasing in P^D until customer participation constraint (2.16) binds. Hence, in equilibrium, naïve customers pay for the premium product

$$P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right). \tag{2.17}$$

Given that $V'(q^*) > 0$ for $0 < q^* < \frac{1}{2}$, it is clear that in equilibrium naïve customers pay more for product A than alert customers.

The equilibrium cutoff with naïve customers is

$$q_N^{*,EQ} = \frac{1}{2} - \frac{\Delta_u V(1/2)}{2\rho}, \quad \text{for } \Delta_u V\left(\frac{1}{2}\right) < \rho, \quad \text{and}$$
$$q_N^{*,EQ} = 0, \quad \text{for } \Delta_u V\left(\frac{1}{2}\right) \geq \rho. \tag{2.18}$$

<u>Proposition 2.2</u>. *Provided that participation constraint (2.14) is satisfied, the outcome with naïve customers is that the financial institutions uses advice cutoff* $q_N^{*,EQ}$ *in* (2.18) *and naïve customers pay* $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$ *for product A. As* $P_N^{D,EQ} > P_A^{D,EQ}$, *financial advice is more biased than with alert consumers, and becomes completely uninformative when* $P_N^{D,EQ} \geq \rho$.

Appendix 2.A.2 gives a proof of Proposition 2.2. Because naïve consumers erroneously expect unbiased advice, they are worse off than without advice. Indeed, substituting (2.17) and (2.18) into (2.6) gives

$$\widetilde{CS}^{D,EQ} = u_l + \Delta_u\left[1 - UC\left(q_N^{*,EQ}\right)\right] - \Delta_u \int_{1/2}^{1} (2q-1)dG(q) < u_0,$$

since $\left[1 - UC\left(q_N^{*,EQ}\right)\right] - \int_{1/2}^{1}(2q-1)dG(q) < \frac{1}{2}$. As long as advice remains informative, total welfare still increases with advice, though. Indeed, total surplus in (2.8) exceeds u_0 when the probability of suitable choice is more than 50%, which is the case with informative advice.

2.4.3 Both alert and naïve customers

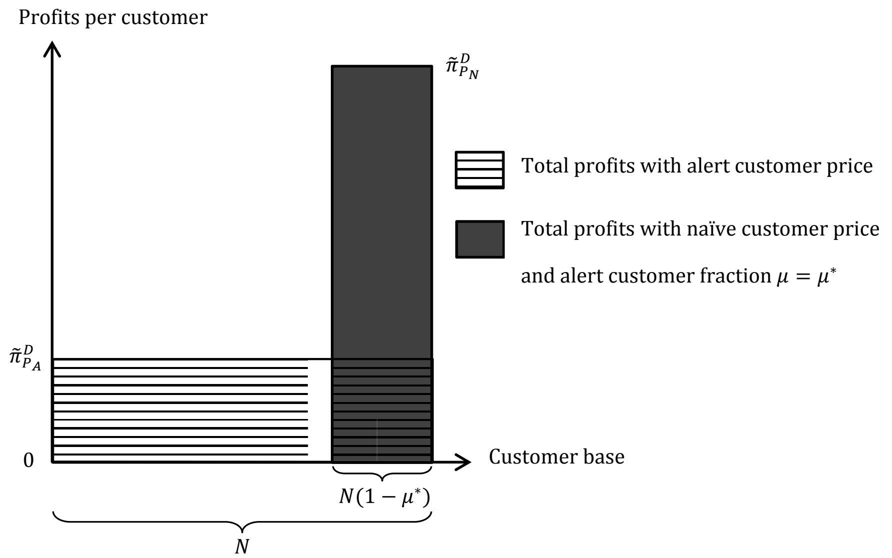
Thus far we have analyzed markets with either alert or naïve customers. Suppose now that a fraction μ of the customers are alert, and $1-\mu$ customers are naïve, and that the financial institution cannot engage in price discrimination. No price discrimination implies that in $t=1$ the firm has to set a single direct sales price P^D for all customers. There are two likely alternatives. The first option is to try to sell to all N customers with advice, which is possible at the alert customer price as characterized in Proposition 2.1. Naïve customers are certainly willing to obtain advice at that price, as they believe that the expected payoff exceeds the payoff from non-advised sales. The second option is to give advice only to naïve customers, which happens when P^D is set at the naïve customer price $P_N^{D,EQ}$. Given Proposition 2.2, we know that this second option leads to a higher premium product price. Total profits may be lower, however, because alert customer won't pay $P_N^{D,EQ} > P_A^{D,EQ}$ for A.

Equilibrium. Denote per customer profits with the alert and naïve customer price $\tilde{\pi}_{P_A}^D$ and $\tilde{\pi}_{P_N}^D$, respectively. To guarantee firm participation, assume $\tilde{\pi}_{P_A}^D > 0$, which implies $\tilde{\pi}_{P_N}^D > 0$. Whether the alert or the naïve customer price maximizes total profits depends on the fractions of alert and naïve customers in the population. Define cutoff $0 < \mu^* < 1$, for which it holds that $\tilde{\pi}_{P_A}^D = (1-\mu^*)\tilde{\pi}_{P_N}^D$. When the fraction of alert customers becomes sufficiently large, defined by $\mu \geq \mu^*$, the financial institution finds it optimal to sell A at the alert customer price, provided that expected profits are non-negative of course. In this case the fraction of non-advised sales $\varepsilon = 0$, and total profits are, by (2.2), $\pi = N\tilde{\pi}_{P_A}^D$ (we still fix $\gamma = 0$). On the other hand, when $\mu < \mu^*$, only naïve customers obtain advice. In this equilibrium, $\varepsilon = \mu$ and total profits are $\pi = N(1-\mu)\tilde{\pi}_{P_N}^D$. Figure 2.1 shows graphically that when the fraction of alert customers $\mu = \mu^*$, total profits with the alert customer price equal total profits with the naïve customer price. Note that the naïve customer price is also chosen if the alert customer price leads to a negative profit margin whereas the naïve customer price is still profitable.

Proposition 2.3. *With both alert and naive customers, and on condition that direct advice is profitable ($\tilde{\pi}_{P_A}^D > 0$), there are two possible equilibrium outcomes. If the fraction of alert customers is relatively large, that is, if $\mu \geq \mu^*$ for a cutoff $0 < \mu^* < 1$, then the equilibrium outcome is as characterized in Proposition 2.1. When $\mu < \mu^*$, however, the financial institution sets the premium product price as defined in Proposition 2.2. In that case only naïve customers will obtain advice.*

Interestingly, when a sufficiently large fraction of the population is alert, naïve customers are protected from exploitation and achieve consumer surplus u_0. In the next section we show that with two distribution channels, naïve customers are not protected by their alert counterparts.

Figure 2.1. Total profits with alert and naïve customer price



Note: $\tilde{\pi}^D_{P_A}$ and $\tilde{\pi}^D_{P_N}$ are per customers profits with product A priced at the alert and naïve customer price, respectively.

2.5 TWO SOURCES OF FINANCIAL ADVICE

While Section 2.4 has looked exclusively at direct sales, this section explores the market outcome when intermediated sales are also allowed for. Recall that the intermediary does not produce financial products, yet advises on product suitability and distributes products to customers. The intermediary can offer product A at price P^I, which may differ from direct advised sales price P^D. The adviser has the same information on product suitability as the financial institution, and also gets regulatory fine ρ in case of unsuitable advice. Note that when the intermediary gives advice, the financial institution cannot be fined, and vice versa. As before, information is exogenously given and comes at no costs. The customer base is assumed to be heterogeneous, with a fraction μ alert customers, and a fraction $1-\mu$ naïve customers. Since this chapter is focused on market dynamics when commissions are banned, the adviser is remunerated exclusively through a fee for advice $f \geq 0$.

In Subsection 1 we explore the baseline model, where the financial institution sets its product prices first, and the other players take these prices as given. By contrast, in Section 2.5.2 we assume that the intermediary moves first and gives a take-it-or-leave-it offer to the financial institution.

2.5.1 Baseline model

Customer participation. In $t = 3$, customers decide between the three sales channels: non-advised sales, direct advised sales and intermediated sales. Let us focus first on the decision between intermediary and non-advised sales. With intermediary sales, recall that the advice cutoff is $q^* = \frac{1}{2}$. When customers decide between intermediary advice and no advice, both alert and naïve customers choose the former if

$$u_l + \Delta_u \int_0^{1/2} (1-q)dG(q) + \int_{1/2}^{1} [\Delta_u q - P^I]dG(q) - f \geq u_0,$$

which can be simplified to participation constraint

$$P^I \leq \Delta_u V\left(\tfrac{1}{2}\right) - \frac{f}{1 - G\left(\frac{1}{2}\right)}. \tag{2.19}$$

In $t = 6$, customers decide whether to follow the intermediary's advice or not. The analysis of this decision is similar to the direct advice analyses in the previous section. That is, customers follow the intermediary's advice to buy product A if

$$u_l + \Delta_u \int_{1/2}^{1} q \frac{dG(q)}{1 - G(1/2)} - P^I \geq u_l + \Delta_u \int_{1/2}^{1} (1-q) \frac{dG(q)}{1 - G(1/2)},$$

which simplifies to participation constraint

$$P^I \leq \Delta_u V\left(\tfrac{1}{2}\right). \tag{2.20}$$

Comparing (2.19) and (2.20), and noting that $f \geq 0$, we find that if customers opt for intermediary advice in $t = 3$, they follow advice to buy A in $t = 6$. Similarly, it can be shown that customers always follow advice to buy product B at zero costs (cf. analysis in Section 2.4). Hence, we conclude that intermediary advice is always followed.

The choice between direct advice and no advice has been analyzed in Section 2.4. Recall that for alert customers this choice is determined by participation constraint (2.11); for naïve customers it is determined by (2.16).

What remains to be investigated is the choice between direct and intermediary advice. Naïve customers believe that the financial institution and the intermediary both use advice cutoff $q^* = \frac{1}{2}$. Consequently, their choice between both advisers is solely driven by product prices P^D and P^I, and advisory fee f. Naïve customers choose direct rather than intermediary advice if

$$P^D \leq P^I + \frac{f}{1 - G\left(\frac{1}{2}\right)}. \tag{2.21}$$

So, if the direct advised sales price of A is sufficiently low compared to the intermediated sales price and the advisory fee, naïve customers prefer direct advice, though the advice quality is certainly lower. Whether naïve customers opt for direct advice depends also on constraint (2.16). If this constraint is satisfied, naïve customers prefer direct advice to no advice, and find it optimal to follow the advice given.

Alert customers recognize that advice quality may differ between the direct and intermediary channel. They prefer intermediary to direct advice if (derivations are in Appendix 2.A.4)

$$P^I \leq \frac{P^D[1 - G(q^*)]}{1 - G\left(\frac{1}{2}\right)} - \frac{f}{1 - G\left(\frac{1}{2}\right)} + \Delta_u \int_{q^*}^{\frac{1}{2}} (1 - 2q) \frac{dG(q)}{1 - G\left(\frac{1}{2}\right)}. \tag{2.22}$$

As alert customers realize that the quality of advice may differ between the two types of advisers, constraint (2.22) is more complex than the analogous naïve customer constraint (2.21). Intuitively, when $q^* = 1/2$, the quality of advice is actually the same and, consequently, (2.22) and (2.21) are alike, the only difference being the sign of the inequality. For $q^* < 1/2$, the intermediary offers better advice than the financial institution. Since alert customers realize this, they are willing to pay a higher intermediated sales price P^I than naïve customers, for given P^D and f. This follows from a comparison of (2.21) and (2.22). Naïve customers prefer direct sales when $P^I \geq P^D - \frac{f}{1-G(1/2)}$, while alert customers prefer intermediary advice for some $P^I \geq P^D - \frac{f}{1-G(1/2)}$.

All in all, we have derived five customer participation constraints. Constraint (2.19) applies to both naïve and alert customers; (2.16) and (2.21) are applicable only to naïve customers; and (2.11) and (2.22) are applicable only to alert customers.

Intermediary and financial institution participation. Although the adviser is unbiased, it may still give unsuitable advice as a result of imperfect information. In case of unsuitable choice, the adviser is fined by the regulator, just as the financial institution is fined when its advice is followed by unsuitable choice. In view of that, the adviser's participation constraint is

$$f \geq \rho UC\left(\frac{1}{2}\right). \tag{2.23}$$

A necessary condition for the financial institution to engage in intermediated sales is that the resulting profits are non-negative. Since fines are payable by the intermediary, $P^I \geq 0$ guarantees firm participation. Note that financial institution participation constraint with direct sales is (2.14).

Equilibrium. In $t = 1$, the financial institution sets prices P^D and P^I with the aim to maximize profits π in (2.2). Recall that of the N customers, a fraction γ choose intermediated sales, a fraction ε choose non-advised sales, and all others are advised by the financial institution. Per customer profits $\tilde{\pi}^D$ and $\tilde{\pi}^I$ determine to what extent the financial institution wants to steer customers toward direct respectively intermediated sales.

It is convenient to start with an analysis of $\tilde{\pi}^I$. (2.4) shows that $\tilde{\pi}^I$ increases linearly in intermediated sales price P^I. Given participation constraints (2.19) and (2.23), the intermediated sales price has upper bound $P^{\bar{I}} = \Delta_u V\left(\frac{1}{2}\right) - \frac{\rho UC(1/2)}{1-G(1/2)}$. When $P^I = P^{\bar{I}}$, per customer profits for the intermediary are $\tilde{F} = f - \rho UC\left(\frac{1}{2}\right) = 0$. When $P^I > P^{\bar{I}}$, the intermediary leaves the market and the game is reduced to the model with direct sales only, which we analyzed in Section 2.4.

Substituting $P^I = P^{\bar{I}}$ in (2.4) yields

$$\tilde{\pi}^{\bar{I}} = \Delta_u \left[\frac{1}{2} - UC\left(\frac{1}{2}\right)\right] - \rho UC\left(\frac{1}{2}\right). \tag{2.24}$$

(2.24) gives maximum profits that can be earned per customer via the intermediary channel.

$\tilde{\pi}^D$ is increasing in direct sales price P^D. As discussed in 2.4.3, there are two plausible alternatives for P^D, that is, the alert customer price as characterized in Proposition 2.1 and the higher naïve customer price as characterized in Proposition 2.2. Substituting the alert customer price and advice cutoff into (2.3), per customer profits are

$$\tilde{\pi}^D_{P_A} = \Delta_u \left[\tfrac{1}{2} - UC\left(q_A^{*,EQ}\right)\right] - \rho UC\left(q_A^{*,EQ}\right), \tag{2.25}$$

where $q_A^{*,EQ}$ is the equilibrium cutoff with alert customers, defined in Proposition 2.1. Comparing (2.24) and (2.25), we observe that $\tilde{\pi}^{\bar{I}} > \tilde{\pi}^D_{P_A}$, since $q_A^{*,EQ} < \frac{1}{2}$ and $UC(q^*)$ is minimalized for $q^* = \frac{1}{2}$. Accordingly, the financial institution prefers intermediated sales with price $P^{\bar{I}}$ to direct sales with alert customer price $P_A^{D,EQ}$.

The alternative direct sales price of product A is the higher naïve customer price $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$. With $P^D = P_N^{D,EQ}$, per customer profits are

$$\tilde{\pi}^D_{P_N} = \Delta_u \left[\tfrac{1}{2} - UC\left(\tfrac{1}{2}\right)\right]\Gamma - \rho UC\left(q_N^{*,EQ}\right), \tag{2.26}$$

where $\Gamma = \left[1 - \mathrm{G}\left(q_N^{*,EQ}\right)\right]/\left[1 - \mathrm{G}(1/2)\right] > 1$, and $q_N^{*,EQ}$ is the equilibrium cutoff with naïve customers as in Proposition 2.2. From (2.24) and (2.26), it is not directly clear how $\tilde{\pi}^{\bar{I}}$ and $\tilde{\pi}^D_{P_N}$ compare. While direct sales have a higher gross profit margin, they also lead to more unsuitable choice and thus to higher regulatory fines. Nonetheless, it can be shown that $\tilde{\pi}^{\bar{I}} < \tilde{\pi}^D_{P_N}$ for any $q_N^{*,EQ} \in [0, \frac{1}{2})$. See Appendix 2.A.4 for the proof.

Taking into account that $\tilde{\pi}^D_{P_N} > \tilde{\pi}^{\bar{I}} > \tilde{\pi}^D_{P_A}$, $P^{D,EQ} = P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$ as in (2.17). The equilibrium intermediated sales price of product A is

$$P^{I,EQ} = \Delta_u V\left(\tfrac{1}{2}\right) - \frac{\rho UC\left(\frac{1}{2}\right)}{\left[1 - G\left(\frac{1}{2}\right)\right]}, \quad \text{for } \Delta_u V\left(\tfrac{1}{2}\right) \geq \frac{\rho UC\left(\frac{1}{2}\right)}{\left[1 - G\left(\frac{1}{2}\right)\right]}. \tag{2.27}$$

Note that when $\Delta_u V\left(\frac{1}{2}\right) < \frac{\rho UC(1/2)}{[1-G(1/2)]}$, $\tilde{\pi}^{\bar{I}} < 0$ and the financial institution does not offer its products via the intermediary ($\gamma = 0$). Conversely, when $\tilde{\pi}^{\bar{I}} \geq 0$, both direct and intermediary advice are attractive, since this implies $\tilde{\pi}^D_{P_N} > 0$. With $P^{I,EQ}$ as in (2.27), the intermediary sets $f^{EQ} = \rho UC\left(\frac{1}{2}\right)$ in $t = 2$, as this is the maximum fee it can charge to its customers. In this equilibrium, by (2.16), (2.19) and (2.21) naïve customers choose direct advice; by (2.11), (2.19) and (2.22), alert customers opt for intermediary advice. Note that constraint (2.22) is not binding in equilibrium. P^I cannot be higher than $P^{I,EQ}$, however, as alert customers would then prefer no advice to intermediary advice by (2.19).

Proposition 2.4. W*ith both direct and intermediary advice, and on condition that* $\tilde{\pi}^{\bar{I}}$ *in (2.24) is non-negative, a unique separating equilibrium exists. In this equilibrium, alert customers receive unbiased advice from the intermediary in exchange for fee* $f^{EQ} = \rho UC\left(\frac{1}{2}\right)$ *and pay* $P^{I,EQ}$ *for premium product A, as characterized in (2.27). By contrast, naïve customers obtain biased financial advice from the financial institution and pay* $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$ *for product A.*

The equilibrium in Proposition 2.4 is incentive compatible since neither alert nor naïve customers have an incentive to switch distribution channel. All customers have an *expected* surplus u_0, yet naïve customers *actual* surplus is lower, and they would therefore be better off with intermediary advice. Naïve customers do not choose intermediary advice, however, because they overestimate the quality of direct advice. As a result, overall consumer surplus is below Nu_0, irrespective of the fraction of alert customers μ. By contrast, the financial institution is possibly better off with two distribution channels. It earns total profits $\pi = N\left[(1-\mu)\,\tilde{\pi}_{P_N}^{D} + \mu\tilde{\pi}^{\bar{I}}\right]$, which exceed profits without intermediated sales when $\tilde{\pi}^{\bar{I}} > 0$. There are two reasons for this. First, two distribution channels enable the financial institution to offer two contracts and (second-degree) price discriminate between alert and naïve customers. Second, with an intermediary, the financial institution is able to extract more surplus from alert customers. The intuition is that when commissions are banned, intermediaries are able to credibly promise sound financial advice, which benefits the financial institution.

How does this equilibrium compare to Inderst and Ottaviani's (2012b) outcome with a heterogeneous customer base? Recall that Inderst and Ottaviani investigate intermediary advice, where the intermediary can be remunerated either indirectly, through a contingent commission, or directly, through a fee for advice. When indirect price discrimination is possible, the authors find a market equilibrium that is identical to Proposition 2.4 in terms of advice quality, and thus overall welfare. Alert customers receive balanced advice ($q^* = \frac{1}{2}$), whereas advice to naïve customers is tilted towards premium product A ($q^* = q_N^{*,EQ}$). When price discrimination is impossible though, the outcome of Inderst and Ottaviani's baseline model depends on the fraction of alert customers (cf. our Proposition 2.3). When the fraction of alert customers is sufficiently large, all customers receive sound advice. When there are a lot of naïve customers, though, alert customers acquire no advice, whereas naïve customers are exploited ($q^* = q_N^{*,EQ}$).

2.5.2 Robustness exercise: the intermediary moves first

We conclude Section 2.5 with a variation of the baseline model. In this variation, the game starts with the intermediary setting its advisory fee *f* in $t = 1$. The financial institution takes *f* as given and sets its prices P^D and P^I in $t = 2$. The rest of the game remains unchanged.

Equilibrium. The intermediary sets *f* as high as possible and never below $\rho UC\left(\frac{1}{2}\right)$, as that would imply an expected loss. By participation constraint (2.19) we have that an increase in *f* decreases the intermediated sales price P^I that intermediary customers are willing to pay. However, the intermediary does not only need participation of the customers but also of the financial institution. When it is profitable to do so, the financial institution may stop intermediated sales altogether and sell its products only directly. Choosing between a multi-channel distribution strategy and direct sales only, the financial institution prefers the former when

$$(1-\mu)\tilde{\pi}^D_{P_N} + \mu\tilde{\pi}^I \geq \tilde{\pi}^D_{P_A}, \tag{2.28}$$

with $\tilde{\pi}^D_{P_N}$, $\tilde{\pi}^I$ and $\tilde{\pi}^D_{P_A}$ as in (2.26), (2.4), and (2.25), respectively.

Recall from Section 2.4.3 that with a heterogeneous customer base and $\tilde{\pi}^D_{P_A} > 0$, there exists a cutoff value μ^* for which $(1-\mu^*)\tilde{\pi}^D_{P_N} = \tilde{\pi}^D_{P_A}$. Hence, if the fraction of alert customers is small, i.e. $\mu \leq \mu^*$, it follows from (2.28) that the financial institution prefers to have both direct and intermediated sales as long as $\tilde{\pi}^I \geq 0$. Taking this into account, the intermediary sets $f^{EQ} = \Delta_u \int_{1/2}^{1}(2q-1)dG(q)$ in $t = 1$. This leaves zero intermediated sales profits for the financial institution. The intermediary has positive earnings, since $\tilde{\pi}^D_{P_A} > 0$ implies $\tilde{\pi}^{\bar{I}} > 0$ and thus $\Delta_u \int_{1/2}^{1}(2q-1)dG(q) > \rho UC\left(\frac{1}{2}\right)$. Equilibrium prices of premium product *A* are $P^{I,EQ} = 0$ and $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$.

When the fraction of alert customers is large, however, intermediated sales profits of the financial institution cannot be pushed to zero by the intermediary. Indeed, with $\mu > \mu^*$ and $\tilde{\pi}^D_{P_A} > 0$, (2.28) implies that the financial institution will only consider a multi-channel strategy when $\tilde{\pi}^I > 0$. In Appendix 2.A.5 we show that in this equilibrium, $P^{I,EQ} > 0, \rho UC\left(\frac{1}{2}\right) < f^{EQ} < \Delta_u \int_{1/2}^{1}(2q-1)dG(q)$, and the direct advised sales price $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$.

<u>Proposition 2.5</u>. *When the intermediary provides a take-it-or-leave-it offer to the financial institution, the source and quality of advice are the same as in the outcome of the baseline model, characterized in Proposition 2.4. The intermediary product price and advisory fee are different, however, and depend on the fraction of alert customers μ. On condition that* $\tilde{\pi}^D_{P_A} > 0$*, there exists*

cutoff $0 < \mu^* < 1$. *When* $\mu \leq \mu^*$, $P^{I,EQ} = 0$ *and* $f^{EQ} = \Delta_u \int_{1/2}^{1} (2q-1) dG(q)$. *By contrast, when* $\mu > \mu^*$, $P^{I,EQ} > 0$ *and* $\rho UC\left(\frac{1}{2}\right) < f^{EQ} < \Delta_u \int_{1/2}^{1} (2q-1) dG(q)$.

Proposition 2.5 implies that the main results of this chapter are robust to changing the sequence of the first two stages of the game. The distribution of intermediated sales profits does change, however. When the intermediary moves first and the fraction of alert customers is small, all intermediated sales profits are pocketed by the intermediary. By contrast, when there are many alert customers and $\tilde{\pi}_{P_A}^{D} > 0$, the financial institution also earns intermediated sales profits, even though the intermediary moves first. The intuition here is that with many alert customers, the financial has a credible threat to bypass the intermediary. As a result, the intermediary rationally shares intermediated sales profits with the financial institution.

2.6 POLICY IMPLICATIONS

Inderst and Ottaviani (2012b) explain that there is an economic rationale for commission bans when customers do not adequately take into account the incentives behind advice. We have shown in Section 2.5 that such unawareness of advice incentives actually makes commission bans ineffective. When commissions are banned, financial institutions have an incentive to by-pass intermediaries and transact with naïve customers directly. To correct the potential market failure of biased direct advice, an obvious - though admittedly radical - intervention would be to prohibit direct advice altogether.

A more gradual way to mitigate the risk of advice bias would be to increase the regulatory penalty for unsuitable advice. With commissions banned, our analysis suggests that policy intervention should focus on advice by financial institutions. By (2.1) we have that the quality of direct advice increases in penalty ρ. In the limit, direct advice may become fully informative, yet it is more likely that financial institutions stop providing advice altogether, as the regulatory fines become prohibitively high. In addition, to increase the probability that misconduct is quickly detected, supervisors could use so-called mystery shoppers. Mystery shoppers act as potential customers and allow supervisors to obtain a detailed insight in how financial firms treat their consumers (FSA 2006b).

Policy intervention could also focus on customers' awareness of conflicts of interests inherent in direct financial advice. Specifically, product advice by financial institutions may need to come with a health warning that the advice given is not independent. To be effective, such a warning needs to turn naïve customers into alert customers, which could be assessed empirically. Research in this area could be initiated by the conduct-of-business supervisors in the countries with commission bans, i.e. the Australian Securities and Investments Commission, the Netherlands Authority for the Financial Markets, and the new Financial Conduct Authority in the UK. Note that transparency also has potential disadvantages. For example, Cain, Loewenstein, and Moore (2005) provide evidence that the disclosure of conflicts of interest may act as a moral-license for self-interested behavior.

2.7 CONCLUSION

There is increasing awareness among policy-makers that commissions-based remuneration of financial advisers leads to biases in financial advice. To improve the quality of financial advice, several large countries have recently adopted laws to ban commission-based remuneration. This chapter explores the incentives behind financial advice when commissions are banned. To this end, we develop a theoretical framework that builds on the work by Inderst and Ottaviani (2012b). In our model, both the financial adviser and the financial institution can give product suitability advice, whereas the original framework only allows for intermediary advice. Since both types of advisers have the same informational advantage vis-à-vis the customers, they are both in a position to increase the probability of suitable choice and, therewith, to increase total surplus in the market.

Our main finding are as follows. With a ban on commissions, the financial institution finds it optimal to offer its products both directly and via an intermediary. Customers who are alert about the incentives behind advice realize that direct advice by the financial institution is of lower quality than intermediary advice. Accordingly, alert customers have a greater willingness to pay for financial products that are advised by an independent financial adviser. Naïve customers, on the other hand, are under the impression that advice is always balanced. The financial institution can benefit from this naïveté through its product pricing. Specifically, it can set its prices in such a way that naïve customers prefer direct advice to intermediary advice. Consequently, in equilibrium, alert customers obtain balanced advice from the intermediary, whereas naïve customers receive advice directly from the financial institution, which is tilted towards the most profitable product. The market outcome thus depends crucially on customer alertness about the incentives behind advice. When a significant fraction of customers is in fact naïve about advice quality, there may be a case for policy intervention. To correct potential market failure, policymakers could consider higher regulatory fines for unsuitable advice and disclosure of conflicts of interests. The latter policy option has the drawback that it may have a permissive effect on advisers for self-interested behavior, potentially making matters worse not better (Cain, Loewenstein, and Moore 2005).

Future work could enrich our analysis by introducing competition between financial institutions and intermediaries. Inderst and Ottaviani (2012b) go into the effect of competition on financial advice. They show that competition does not only lead to lower product prices, it also increases the quality of advice that is given to naïve customers. Our intuition is that competition has the same effect in our model, that is, it will drive prices closer to marginal costs and therewith reduce the incentive to tilt advice to profitable products. Another possible extension of this research is to endogenize the quality of information, where advisers incur costs to obtain information. We leave this for future research.

APPENDIX 2.A PROOFS

2.A.1 Proposition 2.1

We build this proof on the result from Section 2.4.1 that when customers choose direct advice, i.e. when (2.13) holds, they find it optimal to follow the advice that is given. Accordingly, total profits π are in (2.2), with $\tilde{\pi}^D$ as in (2.3). The financial institution sets its price P^D with the aim to maximize π. Since $\gamma = 0, \pi = N(1-\varepsilon)\tilde{\pi}^D$. Given that Proposition 2.1 is under the condition that (2.14) holds, we have that $\tilde{\pi}^D \geq 0$. Accordingly, the financial institution prefers advised sales to non-advised sales ($\varepsilon = 0$). Now, for interior q^*, the first-order derivative $\frac{\partial \tilde{\pi}^D}{\partial P^D} = [1 - G(q^*)] + g(q^*)\frac{P^D}{2\rho} + g(q^*)\left(q^* - \frac{1}{2}\right) = [1 - G(q^*)]$, where we have used (2.1). For corner solution $q^* = 0$, $\frac{\partial \tilde{\pi}^D}{\partial P^D} = 1$. Hence, $\tilde{\pi}^D$ is monotonically increasing in P^D for $0 \leq q^* < 1$. Consequently, in equilibrium, the monopolist finds it optimal to set the premium product price equal to alert customers' maximum willingness to pay with advice, by (2.11), that is $P_A^{D,EQ} = \Delta_u V(q_A^{*,EQ})$.

Without further specification of function $V(q^*)$, we can obtain ranges for $P_A^{D,EQ}$ and consequently $q_A^{*,EQ}$. First note that $P_A^{D,EQ} < \rho$ and $q_A^{*,EQ} > 0$, since a price $P^D \geq \rho > 0$ leads by (2.1) to $q^* = 0$ and zero willingness to pay, i.e. $\Delta_u V(0) = 0$. Second note that $P_A^{D,EQ} > 0$ and $q_A^{*,EQ} < \frac{1}{2}$, since $P^D = 0$ is not a possible equilibrium outcome, as this implies $q^* = \frac{1}{2}$ and $\Delta_u V\left(\frac{1}{2}\right) > 0$. Hence $0 < P_A^{D,EQ} < \rho$ and $0 < q_A^{*,EQ} < \frac{1}{2}$.

For interior $q_A^{*,EQ}$, we have by (1) that $P_A^{D,EQ} = \rho\left(1 - 2q_A^{*,EQ}\right)$, or $\Delta_u V\left(q_A^{*,EQ}\right) = \rho\left(1 - 2q_A^{*,EQ}\right)$. It remains to be shown that $\Delta_u V\left(q_A^{*,EQ}\right) = \rho\left(1 - 2q_A^{*,EQ}\right)$ has a (unique) solution. Given the above, it is sufficient to show that $V'(q^*) > 0$ for every $q^* \in \left(0, \frac{1}{2}\right)$. Differentiating under the integral sign, we have $V'(q^*) = \frac{(1-2q^*)g(q^*)}{1-G(q^*)} + \int_{q^*}^{1} \frac{(2q-1)][g(q)]^2}{[1-G(q^*)]^2} dq$. As $g(q^*) > 0$ and $\int_{q^*}^{1} \frac{(2q-1)][g(q)]^2}{[1-G(q^*)]^2} dq > 0$ for $q^* \in \left(0, \frac{1}{2}\right)$, $V'(q^*) > 0$ for every $q^* \in (0, \frac{1}{2})$. Q.E.D.

2.A.2 Proposition 2.2

It is shown in Section 2.4.2 that as long as customer participation (2.16) is satisfied, naïve customers prefer advised sales to non-advised sales, and also follow the advice given in $t = 6$. Consequently, profits are $\pi = N(1-\varepsilon)\tilde{\pi}^D$, with $\tilde{\pi}^D$ as in (2.3). Under the condition that firm participation constraint (2.14) holds, the financial institution also prefers advised sales to non-advised sales and $\pi = N\tilde{\pi}^D$ in equilibrium. From 2.A.1 we have that $\tilde{\pi}^D$ is strictly increasing in price P^D. Hence, the equilibrium price of product A is equal to naïve customers' maximum willingness to pay, i.e. $P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$. Since $V'(q^*) > 0$ for $q^* \in (0, \frac{1}{2})$, we conclude that $P_N^{D,EQ} > P_A^{D,EQ}$, as $q_A^{*,EQ} < \frac{1}{2}$.

Now, for $P_N^{D,EQ} < \rho$, the equilibrium cutoff is, by (2.1), $q_N^{*,EQ} = \frac{1}{2} - \frac{P_N^{D,EQ}}{2\rho} > 0$; when $P_N^{D,EQ} \geq \rho$, however, $q^* = 0$ and product A is always advised. Q.E.D.

2.A.3 Proposition 2.3

With heterogeneous customers, the equilibrium premium product price is either as described in Proposition 2.1 or as described in Proposition 2.2. There are $N\mu$ alert customers and $N(1-\mu)$ naïve customers. Total profits with the naïve customer price are $N(1-\mu)\tilde{\pi}_{P_N}^D$. At the alert customer price, total profits are $N\tilde{\pi}_{P_A}^D$. Recall that the naïve customer price is higher than the alert customer price and that $\tilde{\pi}^D$ is increasing in price P^D, thus $\tilde{\pi}_{P_N}^D > \tilde{\pi}_{P_A}^D$. Hence, on condition that $\tilde{\pi}_{P_A}^D > 0$, there exists a cutoff $0 < \mu^* < 1$ for which $N(1-\mu^*)\tilde{\pi}_{P_N}^D = N\tilde{\pi}_{P_A}^D$. When $\mu \geq \mu^*$ all customers pay the alert customer price. If, however, $\mu < \mu^*$, total profits from selling only to naïve customers, i.e. $N(1-\mu)\tilde{\pi}_{P_N}^D$, are larger than total profits from selling to all customers at the alert customer price, i.e. $N\tilde{\pi}_{P_A}^D$. Q.E.D.

2.A.4 Proposition 2.4

Again, we use the result in the text that when customers opt for advice, they follow the advice given. This holds both for direct and intermediary advice. When advice is followed, the financial institution has profits $\pi = N[(1-\gamma-\varepsilon)\tilde{\pi}^D + \gamma\tilde{\pi}^I]$, with $\tilde{\pi}^D$ and $\tilde{\pi}^I$ in (2.3) and (2.4), respectively. The financial institution sets premium product prices P^D and P^I to maximize π. Since Proposition 2.4 is on condition that providing advice is profitable, that is, (2.14) and (2.23) are satisfied and $P^I \geq 0$, profit maximization excludes non-advised sales in equilibrium ($\varepsilon = 0$).

The financial institution sets P^I as a take-it-or-leave-it offer to the intermediary in $t = 1$. Provided that a non-negative intermediary sales price of product A is possible, P^I will be set at a level that leaves the intermediary with zero profits, or, $\tilde{F} = 0$. By (2.5) this implies $f = \rho UC(\frac{1}{2})$. Combining $f = \rho UC(\frac{1}{2})$ and (2.19) we have $P^{\bar{I}} = \Delta_u V(\frac{1}{2}) - \frac{\rho UC(1/2)}{[1-G(1/2)]}$. Substituting $P^{\bar{I}}$ for P^I in (2.4), we derive $\tilde{\pi}^{\bar{I}} = \Delta_u \int_{1/2}^{1}(2q-1)dG(q) - \rho UC(\frac{1}{2})$. Noting that $UC\left(\frac{1}{2}\right) = \int_0^{1/2} q dG(q) + \int_{1/2}^{1}(1-q)dG(q) = \frac{1}{2} - \int_{1/2}^{1}(2q-1)dG(q)$, we obtain $\tilde{\pi}^{\bar{I}} = \Delta_u[\frac{1}{2} - UC(\frac{1}{2})] - \rho UC(\frac{1}{2})$.

When deciding between direct and intermediated sales, the institution compares per customer profits $\tilde{\pi}^D$ and $\tilde{\pi}^{\bar{I}}$. From 2.A.3 we have that $\tilde{\pi}^D$ is either $\tilde{\pi}^D_{P_N}$ or $\tilde{\pi}^D_{P_A}$. Substituting $P_A^{D,EQ}$ and $q_A^{*,EQ}$ from 2.A.1 into (2.3) gives $\tilde{\pi}^D_{P_A} = \Delta_u \int_{q_A^{*,EQ}}^{1}(2q-1)dG(q) - \rho UC(q_A^{*,EQ}) = \Delta_u[\frac{1}{2} - UC(q_A^{*,EQ})] - \rho UC(q_A^{*,EQ}) < \tilde{\pi}^{\bar{I}}$. So the financial institution prefers intermediated sales with $P^I = P^{\bar{I}}$ to direct sales with $P^D = P_A^{D,EQ}$.

Substituting $P_N^{D,EQ}$ and $q_N^{*,EQ}$ from 2.A.2 into (2.2) gives $\tilde{\pi}^D_{P_N} = \Delta_u V(\frac{1}{2})[1 - \mathrm{G}(q_N^{*,EQ})] - \rho UC(q_N^{*,EQ})$. Comparing $\tilde{\pi}^D_{P_N}$ to $\tilde{\pi}^{\bar{I}}$, the financial institution prefers to sell to naïve customers directly if $\Delta_u V(\frac{1}{2})[1 - \mathrm{G}(\frac{1}{2})] - \rho UC(\frac{1}{2}) < \Delta_u V(\frac{1}{2})[1 - \mathrm{G}(q_N^{*,EQ})] - \rho UC(q_N^{*,EQ})$ with $0 \leq q_N^{*,EQ} < \frac{1}{2}$. Define the function $h(a) := \Delta_u V(\frac{1}{2})[1 - \mathrm{G}(a)] - \rho UC(a)$. Differentiating with respect to a gives, by Leibniz's rule, $h'(a) = \mathrm{g}(a)[\rho(2a-1) - \Delta_u V(\frac{1}{2})]$. Since $\mathrm{g}(a) > 0$, $\rho > 0$ and $\Delta_u V(\frac{1}{2}) > 0$, we know that $h'(a) < 0$ for $a \in [0, \frac{1}{2}]$. Hence, $\tilde{\pi}^{\bar{I}} < \tilde{\pi}^D_{P_N}$ for $q_N^{*,EQ} \in [0, \frac{1}{2})$.

As $\tilde{\pi}^D_{P_N} > \tilde{\pi}^{\bar{I}} > \tilde{\pi}^D_{P_A}$, the financial institution prefers to set $P^D = P_N^{D,EQ}$. When A is priced accordingly, we know from Proposition 2.2 that naïve customers prefer direct advice to no advice. They prefer direct advice to intermediary advice when $\Delta_u \int_0^{1/2}(1-q)dG(q) + \int_{1/2}^{1}[\Delta_u q - P^D]dG(q) \geq \Delta_u \int_0^{1/2}(1-q)dG(q) + \int_{1/2}^{1}[\Delta_u q - P^I]dG(q) - f$, which boils down to constraint (2.21). With $P^I = P^{\bar{I}}$ and $f = \rho UC(\frac{1}{2})$, (21) is satisfied for $P^D = P_N^{D,EQ}$.

As $P^{\bar{I}}$ is derived from (2.19) and $f = \rho UC(\frac{1}{2})$, alert customers prefer intermediary advice to no advice with these prices. They prefer intermediary advice to direct advice when $\Delta_u \int_0^{1/2}(1-q)dG(q) + \int_{1/2}^{1}[\Delta_u q - P^I]dG(q) - f \geq \Delta_u \int_0^{q^*}(1-q)dG(q) + \int_{q^*}^{1}[\Delta_u q - P^D]dG(q)$. Now this can be simplified by noting that with $q^* < \frac{1}{2}$, we have that $\int_0^{1/2}(1-q)dG(q) - \int_0^{q^*}(1-q)dG(q) = \int_{q^*}^{1/2}(1-q)dG(q)$ and $\int_{1/2}^{1} q\ dG(q) - \int_{q^*}^{1} q\ dG(q) = -\int_{q^*}^{1/2} q\ dG(q)$. Using this and rearranging terms gives participation constraint (2.22). With $P^I = P^{\bar{I}}$, $P^D = P_N^{D,EQ}$, and $f = \rho UC\left(\frac{1}{2}\right)$, (2.22) holds, although not with equality.

All in all, on condition that both direct and intermediary advice are profitable, which holds when $\Delta_u V\left(\frac{1}{2}\right) \geq \frac{\rho UC(1/2)}{[1-G(1/2)]}$, in equilibrium $P^{D,EQ} = P_N^{D,EQ} = \Delta_u V\left(\frac{1}{2}\right)$, $P^{I,EQ} = P^{\bar{I}} = \Delta_u V\left(\frac{1}{2}\right) - \frac{\rho UC(1/2)}{[1-G(1/2)]}$, and $f^{EQ} = \rho UC\left(\frac{1}{2}\right)$. With these prices, alert customers opt for intermediary advice, and naïve customers choose direct advice. Q.E.D.

2.A.5 Proposition 2.5

We prove this Proposition through backward induction and start in period $t = 2$. In this period, the financial institution observes f and uses this information when it sets prices P^D and P^I to maximize profits π. From Proposition 2.3 we have that the financial institution has two main alternatives for P^D. First, it can set P^D equal to the alert customer price as characterized in Proposition 2.1, in which case it can directly advise all N customers. Total profits are then $\pi = N\tilde{\pi}_{P_A}^D$, where $\tilde{\pi}_{P_A}^D$ as in (2.25). Second, it can set P^D equal to the higher naïve customer price as in Proposition 2.2, and provide direct advice only to the $N(1-\mu)$ naïve customers. In this case, the financial institution may transact with alert customers via the intermediary. Profit maximization and customer participation constraint (2.19) imply that the financial institution sets $P^I = \Delta_u V\left(\frac{1}{2}\right) - \frac{f}{1-G(1/2)}$, of course provided that $\Delta_u V\left(\frac{1}{2}\right) \geq \frac{f}{1-G(1/2)}$. With both direct and intermediary advice, total profits are $\pi = N[(1-\mu)\tilde{\pi}_{P_N}^D + \mu\tilde{\pi}^I]$, where $\tilde{\pi}_{P_N}^D$ and $\tilde{\pi}^I$ in (2.26) and (2.4), respectively.

The financial institution prefers to have both direct and intermediated sales to only direct sales when $(1-\mu)\tilde{\pi}_{P_N}^D + \mu\tilde{\pi}^I \geq \tilde{\pi}_{P_A}^D$. Provided that $\tilde{\pi}_{P_A}^D > 0$, there exists a cutoff $0 < \mu^* < 1$ for which $(1-\mu^*)\tilde{\pi}_{P_N}^D = \tilde{\pi}_{P_A}^D$. When $\mu \leq \mu^*$, the financial institution prefers to have both direct and intermediated sales, even when $\tilde{\pi}^I = 0$. However, when the fraction of alert customers is large, i.e. $\mu > \mu^*$, the financial institution opts for direct sales at the alert customer price when $\tilde{\pi}^I = 0$.

In $t = 1$, the intermediary sets fee f to maximize profits $F = N\gamma\tilde{F}$, where γ is the fraction of intermediary customers and $\tilde{F}$ are per customer profits. (2.5) shows that $\tilde{F}$ is monotonically increasing in advisory fee f. So, the intermediary sets f as high as possible, that is, without inducing the financial institution to bypass the intermediary altogether ($\gamma = 0$). When $\mu \leq \mu^*$, f can be raised until $\tilde{\pi}^I = 0$. With a binding customer participation constraint (2.19), the equilibrium advisory fee is then $f^{EQ} = \Delta_u \int_{1/2}^{1}(2q-1)dG(q)$, intermediary product price $P^{I,EQ} = 0$, and direct product price $P^D = P_N^{D,EQ}$.

Conversely, when $\mu > \mu^*$ and $\tilde{\pi}^D_{P_A} > 0$, $\tilde{\pi}^I > 0$ in equilibrium. This implies that with many alert customers, the equilibrium advisory fee is $f^{EQ} < \Delta_u \int_{1/2}^{1}(2q-1)dG(q)$ and $P^{I,EQ} > 0$. With $f = \rho UC\left(\frac{1}{2}\right)$, the financial institution earns the maximal amount per intermediary customer, namely $\tilde{\pi}^{\bar{I}}$. The proof of Proposition 2.4 shows that $\tilde{\pi}^{\bar{I}} > \tilde{\pi}^D_{P_A}$. Since the institution prefers to have an intermediary channel when $\tilde{\pi}^I \geq \tilde{\pi}^D_{P_A}$, we have $f^{EQ} > \rho UC\left(\frac{1}{2}\right)$.

Note that, irrespective of the fraction of alert customers μ, the intermediary sets f in such a way that the financial institution prefers to transact with alert customers via the intermediary and with naïve customers directly, using direct advice cutoff $q_N^{*,EQ}$. Hence, in terms of advice quality and source of advice, the outcome of this variation is identical to the outcome of the baseline model as characterized in Proposition 2.4. Q.E.D.

APPENDIX 2.B ILLUSTRATION OF DIRECT ADVICE MODEL

The results of this chapter are general, that is, they hold for any posterior-belief distribution $G(q)$ with density $g(q) > 0$ over $q \in [0,1]$. With the aim of illustrating the model without intermediary involvement ($\gamma = 0$), we sacrifice some generality here and postulate a specific posterior-belief distribution. That is, we assume that the posterior is uniformly distributed over $[0,1]$, i.e. $G(q) = q$. This implies that the financial institution has information about product suitability, yet this information is imperfect. Indeed, with perfect information, the posterior is either $q = 0$ (B is unquestionably the right product) or $q = 1$ (A is unquestionably the right product), both occurring with 50% probability.

With a uniformly-distributed belief and alert customers, A's equilibrium price equals $P_A^{D,EQ} = \Delta_u q_A^{*,EQ}$. Combining this with Equation (2.1) yields possible equilibrium values when all customers are alert:

$$P_A^{D,EQ} = \frac{\Delta_u \rho}{\Delta_u + 2\rho}, \quad \text{and}$$
$$q_A^{*,EQ} = \frac{\rho}{\Delta_u + 2\rho}.$$

These price and cutoff values represent market equilibriums if the firm participation constraint holds. Substituting $q_A^{*,EQ}$ in constraint (2.14), we obtain that firm profits are non-negative for $\Delta_u \geq \rho\sqrt{2}$. When the firm participation constraint does not hold, the financial institution does not provide advice and might even withdraw from the market. Customers then buy either randomly product A or B, or, when the financial institution leaves the market, always B. In both cases consumer surplus and total surplus equal $u_0 = u_l + \frac{1}{2}\Delta_u$.

Using $G(q) = q$ in (2.17) and (2.18) gives possible equilibrium values with naïve customers:

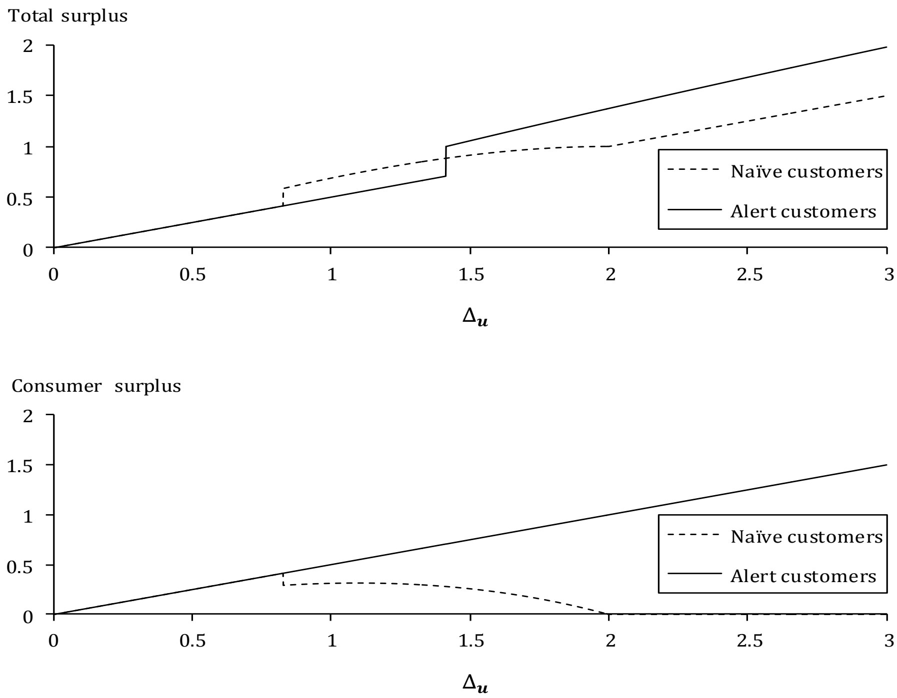
$$P_N^{D,EQ} = \frac{\Delta_u}{2}, \text{ and}$$

$$q_N^{*,EQ} = \frac{1}{2} - \frac{\Delta_u}{4\rho} \quad \text{for } \Delta_u < 2\rho, \text{ and } q^{*,EQ} = 0 \text{ for } \Delta_u \geq 2\rho.$$

The financial institution finds it valuable to provide direct advice to naïve customers if $\Delta_u \geq \rho(2\sqrt{2} - 2)$. Hence, it is more likely to offer advice with naïve customers than with alert customers, as it is profitable to give advice to the former type of customers at lower values of Δ_u (for given ρ). When the constraint is not fulfilled, however, naïve customers either randomize between A and B, or always buy B.

Figure 2.B.1 shows total surplus and consumer surplus for different values of Δ_u.

Figure 2.B.1 Total and consumer surplus with imperfect information



Notes: The following assumptions have been made: posterior belief q follows a uniform distribution, i.e. $G(q) = q$; penalty $\rho = 1$; utility in case of mismatch $u_l = 0$; and customer mass $N = 1$. With $N = 1$, total surplus with direct advice is given by (2.8), and consumer surplus is given by (2.6).

For expositional purposes, we have set $\rho = 1$, $u_l = 0$, and customer mass $N = 1$. Note that when Δ_u increases, match suitability becomes more important to customers, and therewith to the financial institution, as it can charge a higher price for the premium product. The top panel shows that total surplus is monotonically increasing in Δ_u, both with alert and naïve customers. The kinks in the graphs can be traced back to the firm participation constraint. When it is not profitable to provide advice, the monopolist switches to execution only or leaves the market altogether. In both cases total and consumer surplus are $\frac{1}{2}\Delta_u$, and producer surplus is zero.

The lower panel shows that customers are not necessarily better off when Δ_u increases. Alert customers benefit from higher Δ_u, even though advice quality deteriorates. This is because their alertness prevents the financial institution from raising product A's price too much. Naïve customers, however, are strictly worse off with advice. In fact, when $\Delta_u \geq 2$, advice becomes completely uninformative and all consumer surplus is extracted by the financial institution.

Chapter 3

DEDUCTIBLE CHOICE INCENTIVES*

* This chapter is based on Gorter and Schilp (2012).

3.1 INTRODUCTION

In insurance contracts, the deductible is the amount a policyholder has to pay out-of-pocket before the insurer covers the remaining costs. Deductibles are primarily intended to reduce the moral hazard that may arise from insurance coverage. When voluntary, deductibles allow policyholders to fine-tune their insurance coverage. A policyholder that opts for a low deductible is exposed to less risk, yet at the expense of a higher insurance premium. The deductible feature in insurance policies is valuable for research purposes, as it allows for an investigation of small-scale risk taking in the field, rather than in the laboratory. As Loewenstein (1999) argues, field studies have a significant advantage in terms of external validity, i.e. to what extent the results carry over to real-life.

This chapter investigates the incentives behind deductible choice in Dutch universal health insurance. The institutional characteristics of universal health insurance in the Netherlands make it close-to-ideal for studying deductible choice. The so-called basic health insurance policy is mandatory to all Dutch residents. As a result, consumer choice is restricted to choosing a health insurer and a deductible-rebate package with that insurer. Individuals that opt for a higher deductible have the same health plan as everybody else, yet they voluntarily expose themselves to a small amount of financial risk. In 2008, the year which we study, residents of 18 years and older could choose a voluntary deductible from six alternatives, ranging from €0 till €500 (roughly $650) in stages of €100. For a nonzero voluntary deductible policyholders receive a premium rebate that is independent of health status and risk (Van Kleef, Beck, Van de Ven, and Van Vliet 2008). Consequently, asymmetric information is effectively guaranteed and the analysis can be focused on the demand-side. Nonetheless, in the empirical work we verify that rebate differences between insurers do not influence our results.

As premium rebates are independent of risk, we expect adverse selection, that is, more risky policyholder have an incentive to choose a high voluntary deductible. In practice, however, only about 5% of Dutch adults chooses a nonzero voluntary deductible (2008 figure). A potential explanation for this outcome is that the level of the premium rebates is too low, that is, lower than the expected value of the insurance coverage. Another would be that people are too risk-averse, yet since deductibles are small-scale risks compared to lifetime wealth, expected-utility theory predicts that individuals choose their deductible in a risk-neutral fashion. Under risk neutrality, expected value drives decision-making and differences in risk attitudes do not play a role. Arrow (1971) called this the local risk neutrality of expected-utility theory. Rabin (2000) brought the issue of local risk neutrality to the fore again. He showed that if the only reason consumers are risk-averse is diminishing marginal utility of wealth, which is what expected-utility models assume, then consumers should be virtually risk-neutral over stakes in the hundreds or even thousands of U.S. dollars.

Contrary to the prediction of expected-utility theory, we find evidence that Dutch adults are averse to small risks. In fact, using a unique combination of population and survey data, we conclude that healthy individuals' monetary incentive to choose a high voluntary deductible tends to be overshadowed by their risk aversion. In our regressions, risk preferences are both statistically and economically more significant in explaining deductible choice than risk type. This finding is consistent with increasing empirical evidence that people can be significantly averse to small risks. The majority of this evidence comes from experiments that are conducted under controlled laboratory conditions (e.g. Holt and Laury 2002, Barberis, Huang, and Thaler 2006, Harrison and Rutström 2008). Evidence from real market settings is scarcer. There are a few existing field studies, however. Cicchetti and Dubin (1994) study the demand for telephone wiring protection in the U.S, which is surely a small risk with own payments maximized to $55. Cohen and Einav (2007) estimate risk preferences from data on deductible choices in Israeli auto insurance contracts. Sydnor (2010) uses data on deductible choices in U.S. homeowner's insurance to calibrate a bound for the implied level of risk aversion. While each of these studies uses the expected-utility framework to estimate risk preferences over modest risk, the results differ. Both Cicchetti and Dubin (1994) and Cohen and Einav (2007) report relatively low levels of risk aversion, while Sydnor (2010) finds that under expected-utility theory only extreme measures of risk aversion can rationalize the choice for costly low deductibles.

Our contribution to the above-mentioned empirical literature is threefold. First, while related studies investigate deductible choice of policyholders in a particular insurance market, we investigate a universal insurance market, which covers the majority of health risks of *all* Dutch residents. Hence, our results are generalizable to the Dutch population. Second, we combine unique population data with representative survey data, which allows us to give a relatively detailed evaluation of deductible choice behavior in the Dutch universal health insurance market. The population data include information on deductible choice, premium rebates and out-of-pocket expenditures. The survey data cover members of the CentERpanel; a panel of Dutch households that is designed to be representative of the Dutch population. We refer to Teppa and Vis (2012) for a comprehensive description of the CentERpanel.

An important feature of our survey data set is that it includes good proxies for both risk type and risk preferences. This brings us to our third contribution, namely that we use direct measures of risk preferences to explain deductible choice. Related studies infer risk tolerance parameters by assuming expected-utility-of-wealth maximization and a specific utility function. Hence, this work also fits within a growing stream of research that measures risk preferences from survey data and relates these direct measures of risk tolerance to actual consumer behavior (e.g. Barsky, Juster, Kimball, and Shapiro 1997, Guiso and Paiella 2005, 2008, Cutler, Finkelstein, and McGarry 2008).

The rest of the chapter is organized as follows. Section 3.2 provides an overview of the basic health insurance market in the Netherlands and discusses deductible choice and expenditure patterns that emerge from population data. These data show that people by and large choose the lowest deductible, though young men have low expected out-of-pocket expenditures and would benefit, in expected value terms, from higher levels of deductibility. Section 3.3 outlines our empirical approach and provides underpinning for our proxies for risk type and risk preferences. Section 3.4 describes the survey data set and gives first indications of a relationship between attitudes towards risk and deductible choice in our sample of Dutch adults. Section 3.5 presents the empirical result. By way of preview, we find that risk preferences are both statistically and quantitatively more significant in explaining deductible choice behavior than risk characteristics, which contrasts with standard expected-utility of wealth theory. Potential explanations outside the canonical utility model are discussed. Section 3.6 concludes.

3.2 MARKET DESCRIPTION AND POPULATION CHOICE DATA

3.2.1 Market description

The implementation of the Health Insurance Act in 2006 significantly changed the Dutch market for health insurance. After decades of price and capacity control by government, the Dutch healthcare system shifted from supply-side regulation to managed competition (Van de Ven and Schut 2008). The aim of this shift was to make healthcare more cost-efficient and improve quality.

In the Dutch health insurance system, residents are obliged to purchase the so-called basic health insurance policy from a private health insurer. Each year, the exact composition of this basic package is determined by government. Generally, it includes care provided in hospitals, by general practitioners (GPs) and specialists, prescription drugs, maternity care, obstetrics, technical aids and dental care for children (Van Kleef et al*.* 2008). Though insurers are free to offer preferred-provider policies, insurers tend to cover all healthcare suppliers.

To curtail healthcare consumption arising from moral hazard, the Dutch government initially arranged for a mandatory no-claim refund and a voluntary deductible. Depending on actual health care expenses, policyholders were refunded between €0 and €255 of their premiums. The voluntary deductible alternatives ranged from €0 till €500 in stages of €100. In 2008, the year of our study, the no-claim refund was replaced by a mandatory deductible of €150. The six voluntary deductible alternatives remained the same.

While insurers are free to set their premiums and premium rebates, price differentiation by risk type is strictly prohibited. Both premiums and premium rebates are community-rated, i.e. all policyholders of a certain insurer pay the same premiums and are offered the same deductible-rebate packages. Hence, premiums and premium rebates are independent of health status and risk. The Dutch health insurance system is truly universal: health insurers are obliged to accept every resident, while residents are obliged to buy basic health insurance. To compensate health insurers for unhealthy and therefore costly pools of policyholders, a Risk-Equalization Fund was set up.[1]

On average, this fund finances 50% of total healthcare expenditures through income-related contributions. Using information on residents' age, gender, region, source of income, pharmacy-based cost groups and diagnostic-based cost groups, an insurer receives equalization payments from this fund (Van Kleef et al*.* 2008). For instance, an insurer with an above average fraction of elderly people in its portfolio is compensated accordingly. About 45% of total health care costs are financed through insurance premiums. Note that children up to the age of 18 are exempted from paying premiums and do not have a deductible, neither mandatory nor voluntary. Medical care of children is financed by the Dutch government and constitutes about 5% of total health care costs.

3.2.2 Population choice data

To obtain insight into deductible choice in Dutch health insurance, we have obtained population data from Vektis, the information centre established by Dutch health insurers, and from the Dutch healthcare authority (NZA). Parts of these data are summarized in Table 3.1. Shown are the deductible-rebate menus offered in 2008, and the distribution of actual choices made. We consider the weighted-mean rebate data most informative, given that these are adjusted for the number of policyholders that were actually offered these deductible-rebate alternatives by their respective health insurers. On average, policyholders could reduce their annual premiums with €211 by opting for a voluntary deductible of €500. The total deductible, i.e. the sum of the mandatory and voluntary deductible, would then be €650 (€150 plus €500). The 10th and 90th percentile of the rebates offered show that rebates differ between health insurers. Since risk selection by insurers is strictly prohibited, it is unlikely that heterogeneity of rebates is important to our analysis. Nonetheless, in the empirical analysis, we will also investigate the effect of premium-rebate differences between health insurers. The last column of Table 3.1

[1] Switzerland has a similar health insurance system, including voluntary deductibles with community-rated premiums and a risk-equalization scheme. Van Kleef et al. (2008) provide a detailed discussion of the Dutch and the Swiss basic health insurance schemes.

shows that in 2008 only 5% of Dutch adults aged 20 years and over held a voluntary deductible of €100 or more. In 2007 this percentage was even lower (4%, not shown in Table 3.1).

Table 3.1. Deductible-rebate menus offered and choices made in 2008

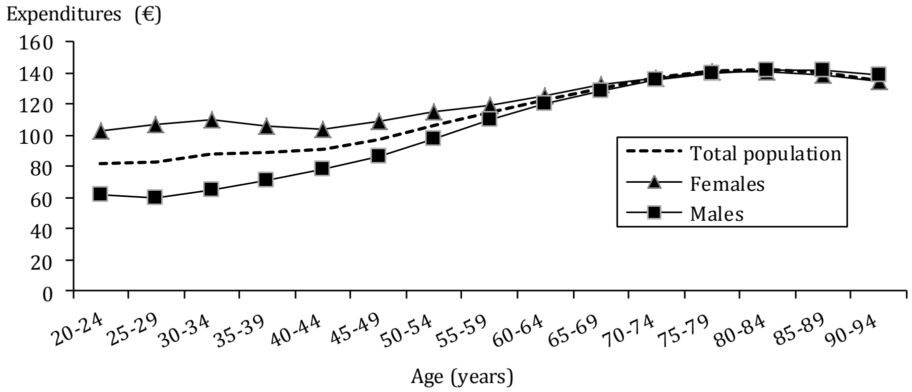
Voluntary deductible (€)	Offered community-rated rebates (€)			Population distribution (adults)	
	Weighted mean*	10th percentile	90th percentile	Number (thousands)	%
0	-	-	-	11,386	94.9
100	47	30	50	182	1.5
200	90	71	104	102	0.9
300	132	100	150	59	0.5
400	172	139	203	17	0.1
500	211	180	250	253	2.1
Total				*12,000*	*100*

* Rebates are weighted by the number of policyholders of the concerning health insurer. *Notes*: Deductible-rebate menus offered by the different health insurers are from the NZA. The distribution of deductible choice is from Vektis

Figure 3.1 presents average out-of-pocket expenditures in 2008, for different age categories and by gender.[2] These data are from Vektis, the healthcare information centre established by Dutch health insurers. The plotted average out-of-pocket expenditures are for the entire Dutch population and thus include individuals with both zero and nonzero voluntary deductibles. However, as the lion's share of Dutch adults chose a voluntary deductible of €0 in 2008, and thus had a total deductible of €150, average out-of-pocket expenditures are fairly low (below €150). The figure shows that for both males and females, out-of-pocket payments increase with age. Moreover, males generally have lower expenditures than females, yet this gender difference decreases with age. Young women have significantly higher health expenditures than young men, primarily because of pregnancy costs. While visits to GPs, obstetrics and maternity care are not subject to out-of-pocket payments, other pregnancy-related healthcare costs are. Up to the deductible amount, pregnant women must pay for laboratory research, medicines and ambulance transport to hospital themselves. Another factor that influences out-of-pocket payments by young women is the contraceptive pill, which was subject to out-of-pocket expenses in 2008 (not in 2007).

[2] Note that out-of-pocket expense data may underestimate actual expenditures since people could decide to pay themselves, for example for medications, without notifying their health insurer. This underestimation of out-of-pocket expenditures is unlikely to be substantial, since many healthcare suppliers send their bills directly to health insurers. Moreover, people have an incentive to claim, since it reduces that year's remaining amount of potential out-of-pocket expenditures. Note that we also have 2007 data on out-of-pocket expenditures, and the relationships with age and gender are qualitatively the same.

Figure 3.1. Average out-of-pocket basic health expenditures Dutch population in 2008



Notes: Data are from Vektis. Averages have been calculated for all Dutch residents in the respective age categories. All possible voluntary deductible levels are thus included. In 2008, 95% of the adult population had a voluntary deductible of €0 (see Table 3.1).

Based on the average expenditure patterns in Figure 3.1, one would expect that young people in general and young males in particular are more inclined to choose a nonzero voluntary deductible. Table 3.2 shows that this is indeed the case, though the differences between males and females, and young and old are fairly small.

Table 3.2. Voluntary deductible choice Dutch population in 2008, by gender and age

Gender	Age (years)	Percentage with voluntary deductible €0	Percentage with voluntary deductible > €0
male	20-29	92	8
male	30-39	91	9
male	40-49	92	8
male	50-59	93	7
male	60-69	96	4
male	70-79	99	1
male	80-89	99	1
female	20-29	95	5
female	30-39	95	5
female	40-49	95	5
female	50-59	96	4
female	60-69	98	2
female	70-79	99	1
female	80-89	99	1

Note: Data are from Vektis.

To assess the attractiveness of choosing a nonzero deductible in expected value terms, average out-of-pocket expenditures are of little use. Fortunately, we do not only have averages of the out-of-pocket expenditure distributions but also deciles, which can be used to this end. Table 3.A.1 in Appendix 3.A gives the out-of-pocket expenditure data we have. These data describe the expenditure patterns of Dutch adults with a €0 voluntary deducible, i.e. 95% of the adult population.

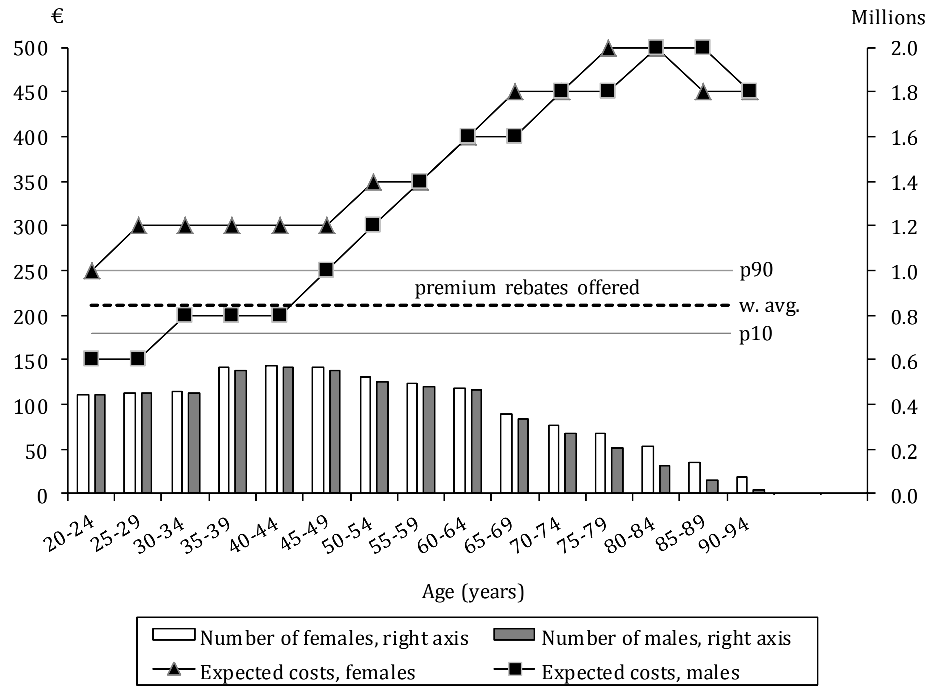
Deciles of the out-of-pocket expenditure distributions can be used to determine an upper bound of the expected additional costs of increasing the voluntary deductible. We speak of *additional* expenditures, since these expenditures come on top of the out-of-pocket expenses under the mandatory deductible of €150. Define decile number *x* as the lowest decile of a certain age-gender expenditure distribution for which actual out-of-pocket expenditures equal the total deductible of €150. For individuals in this age-gender group, an upper bound of the expected additional costs from choosing the maximum deductible is then €500*(1.1-0.1*x*). This method leads to an upper bound of expected out-of-pocket expenditures for two reasons. First, it assumes that the (10-*x*)*10% of the age-gender group that had out-of-pocket expenses equal to the actual deductible of €150 would have had the maximum out-of-pocket expenses of €650 with a voluntary deductible of €500. Second, it assumes that the mass of the expenditures distribution between deciles *x*-1 and *x* is concentrated infinitely close to decile *x*. By making these assumptions, we are able to identify age-gender groups in the Dutch population for which a voluntary deductible of €500 would be attractive in expected value terms, even if the expected additional expenditures are overestimated.

An example may further clarify our procedure. For males aged 40-44 years with an actual voluntary deductible of €0, the 7th decile is the lowest decile for which out-of-pocket expenses are €150 (see Table 3.A.1 in Appendix 3.A). Consequently, we know that at least 60% of the males in this age group had actual out-of-pocket expenditures that were lower than the mandatory deductible of €150. An upper bound of the expected additional expenditures from choosing a voluntary deductible of €500 is then €200 for this population group. The expected additional costs are lower than the average premium rebate of €211 (see Table 3.1), making a voluntary deductible attractive with risk-neutral preferences.

Using the procedure described above, we have calculated upper bounds of the expected additional expenditures from increasing the voluntary deductible from €0 to €500 for different age-gender groups. Figure 3.2 presents the results. For comparison reasons, we have also included information on the range of premium rebates that were offered for a voluntary deductible of €500 as well as the population distribution of males and females over these age cohorts.

From Figure 3.2 we infer that for the majority of the population, premium rebates were probably too low to make a voluntary deductible of €500 attractive. Hence, the low appetite for nonzero voluntary deductibles is at least partly the result of low premium rebates. That being so, a €500 deductible does seem attractive for relatively young men (aged 44 years and under). Their expected additional costs from raising their voluntary deductible with €500 are certainly lower (since we calculated upper bounds) than the average premium rebate of €211. In practice, however, only 8% of the 2.7 million men aged 20-44 years chose a voluntary deductible higher than €0 in 2008 (2007 figures are similar). Note that for male adults of 29 years and younger, the calculated upper bound of expected additional expenditures is even lower (€150), which makes a voluntary deductible of €500 even more attractive. Still, just 8% of the nearly 1 million men aged 20-29 years held a nonzero voluntary deductible (see Table 3.2). Hence, the population data indicate that risk aversion holds young men back to choose a nonzero voluntary deductible.

Figure 3.2. Costs and benefits of increasing voluntary deductible from €0 to €500 in 2008



Notes: Out-of-pocket expenditure data are from Vektis. Data on premium rebates offered are from the NZA. Upper bounds of the expected additional expenditures have been calculated as described in the text.

3.3 EMPIRICAL APPROACH

Modeling deductible choice in an empirical setting can be quite challenging. It requires a detailed understanding of the risks that are insured, the features of the contracts traded and the exact distribution of information between buyers and sellers (Chiappori and Salanié 2008). If insurers have information on their (would-be) policyholders and they are allowed to use this information in their pricing, the estimation methodology should correct for that. While early studies such as

Dahlby (1983, 1992) and Puelz and Snow (1994) did not have full access to insurers' information, recent works such as Cohen (2005) and Saito (2006) are based on all data from the relevant insurance company.

Thanks to the institutional features of universal health insurance in the Netherlands, our empirical approach can be relatively straightforward. As health insurers are obliged to accept every eligible applicant at community-rated premiums, asymmetric information is effectively guaranteed. When information is asymmetrically distributed between insurers and policyholders, risk type is an important potential driver of deductible choice. Indeed, with community-rated premium rebates, both theory and available empirical evidence predict adverse selection, which means that high risk individuals choose a low deductible, and vice versa (literature reviews are provided by Cutler and Zeckhauser 2000, and Cohen and Siegelmann 2010).[3]

While risk selection on the demand side is an important topic in itself, we are primarily interested in the importance of risk preferences in explaining deductible choice behavior. Although expected-utility theory predicts risk neutrality over modest stakes, actual choice behavior suggests differently. To investigate the importance of risk preferences in modest scale risk taking, we estimate the following model of deductible choice:

$$\widetilde{D}_{i,t} = f\left(Risk_{i,t-1}, Risk\ preference_{i,t-1}, X_{i,t-1}, \beta\right) + \varepsilon_{i,t}, \tag{3.1}$$

where the dependent variable $\widetilde{D}_{i,t}$ is a latent variable measuring consumer i's desired level of deductibility in year t. $\widetilde{D}_{i,t}$ is assumed to be a function $f(\cdot)$ of the explanatory variables in vectors $Risk_{i,t-1}$, $Risk\ preference_{i,t-1}$ and $X_{i,t-1}$, and the parameters in vector β. $f(\cdot)$ is linear in β. Since $\widetilde{D}_{i,t}$ is unobserved, we use actual voluntary deductible choice, $D_{i,t}$, instead. Equation (3.1) is estimated both as a probit model, where the dependent is 1 for individuals with a nonzero voluntary deductible, and as an ordered probit model, where the dependent can take one of the six voluntary deductible levels. Disturbance term $\varepsilon_{i,t}$ is (thus) assumed to follow a normal distribution.

The explanatory variables in Equation (3.1) are all lagged one year as deductibles are chosen ex-ante. $Risk_{i,t-1}$ is a vector of risk variables, two of which are age and gender. In Section 3.2 we have shown that there are distinct expenditure differences between males and females, and that out-of-pocket expenditures by and large increase with age. As the gender expenditure

[3] It is worth noting that adverse selection, or a positive correlation between insurance coverage and risk, is not an empirical regularity across insurance markets. Cohen and Siegelmann (2010) review the empirical literature on adverse selection in insurance markets and conclude that "a risk-coverage correlation appears to be a feature of some insurance markets [...] but not of others".

difference is most pronounced at young ages, a dummy for young men (which is 1 for men aged 44 years and under, and 0 otherwise) is also included in some of the regression analyses. As a group, young men have the lowest out-of-pocket expenditures and are therefore expected to be more inclined to choose a nonzero voluntary deductible. Note that gender is also a potential explanatory variable for heterogeneity in risk tolerance levels. Indeed, Barsky et al. (1997) and Hartog, Ferrer-i-Carbonell, and Jonker (2002) find that women are significantly more averse to risk than men. Although we control for differences in risk aversion in our regression analyses, it is conceivable that, to some extent, risk preferences are still picked up by the gender dummy.

Besides age and gender, we use self-assessed health status (SAHS) and number of GP visits as risk proxies. SAHS is a subjective risk measure and is generally regarded to be a good predictor of future health conditions (e.g. Gerdtham, Johannesson, Lundberg, and Isacson 1997, Wagstaff and Van Doorslaer 1994). Moreover, like all subjective measures, SAHS has the advantage of strictly reflecting information known to the consumer. Consequently, unknown aspects of one's health condition – which by definition cannot play a role in deductible choice – are rightly ignored. The SAHS survey question is formulated as follows: 'What is your health like in general?,' with five response categories ranging from 'excellent' to 'poor'. A potential drawback of SAHS is that survey respondents may implicitly assess their health relative to their age category (Buchmueller, Fiebig, Jones, and Savage 2009, Doiron, Jones and Savage 2008). To address this concern, we also interact SAHS with age. GP visits are included as a risk measure, although they are not subject to out-of-pocket payments. The rationale is that a visit to the GP increases the probability of other healthcare use and thus out-of-pocket payments. For example, GPs give prescriptions for medications and refer patients to specialists, both leading to out-of-pocket expenses.

The second vector of explanatory variables in Equation (3.1), $Risk\ preference_{i,t-1}$, contains risk-preference variables. One of these is a direct measure of financial risk tolerance, that is, we use answers to the following statements: 'I am willing to run the risk of losing money if there is also a chance that I will make money.' Respondents can answer this question on a seven point scale, where one means 'completely disagree' and seven means 'completely agree.' This measure of financial risk tolerance is quite fitting, as a higher deductible only involves more financial risk. Another potentially relevant risk-preference variable is wealth. Either with decreasing absolute risk aversion (DARA) preferences or with constant relative risk aversion (CRRA) preferences, the willingness to take risk increases with lifetime wealth. While wealth would be irrelevant under constant absolute risk aversion (CARA) preferences, there is increasing evidence suggesting that absolute risk aversion decreases with wealth (e.g. Guiso and Paiella 2008). Since lifetime wealth can be seen as the sum of accumulated wealth and human capital, we use accumulated financial wealth and annual income – the latter as a measure of human capital – to proxy lifetime wealth.

In addition to proxies of financial risk tolerance and lifetime wealth, $Risk\ preference_{i,t-1}$ contains four indicators of risky behavior. These behavioral variables relate to the smoking and drinking behavior of respondents, their job risk and their holdings of risky financial assets. Smoking and drinking have frequently been used as indicators of risky behavior (e.g. Barsky et al. 1997, Cutler et al. 2008, Doiron et al. 2008). Dummy variable 'smoking' is one for daily smokers; dummy variable 'drinking' is one for individuals with daily alcoholic consumptions in excess of four. One could argue that smokers and drinkers have higher out-of-pocket expenses, and that, therefore, these dummy variables are also relevant risk proxies. Although we are unaware of any formal evidence of this, such measurement bias in these risk preference proxies, if any, would lead to an underestimation of the effect of risk preferences on deductible choice. Job risk is measured by a self-employment dummy variable. Self-employed individuals typically have a riskier income stream than employees (e.g. Friedman 1957, Carroll 1994). Our last risk-preference variable measures the ownership of risky financial assets by the portfolio share of equities (cf. Guiso and Paiella 2006, 2008).

The final vector of explanatory variables in Equation (3.1) is $X_{i,t-1}$, which controls for heterogeneity in, for instance, education and number of children. These and all other variables are described in the next section.

3.4 SURVEY DATA

We use data on individual CentERpanel members that have been collected through internet surveys of CentERdata.[4] The CentERpanel was established in 1991 and consists of over 2,000 households.[5] The panel is an appropriate representation of the Dutch-speaking population in the Netherlands and has been used before by, among others, Van Rooij et al. (2007). The questionnaires are answered at home, so participants do not feel rushed to give an answer and are fully anonymous when answering the questions. Chiang and Krosnick (2010) argue that when compared to telephone interviewing, internet surveys exhibit higher validity and less social desirability response bias. Participants do not receive payment for their participation.

Our survey on deductible choice in basic health insurance was sent to 1,826 panel members that have indicated to be the principal financial decision-maker of their respective households. We asked them to specify their basic health insurer (from a list of 30) in 2008, as well as their

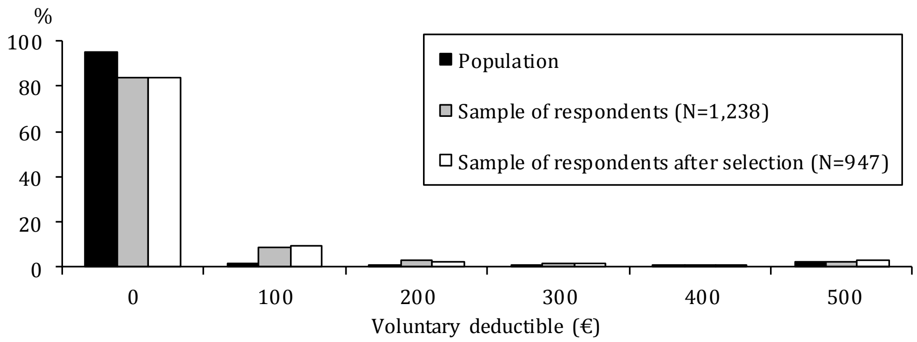
[4] CentERdata forms part of the CentER Group at Tilburg University. See also http://www.uvt.nl/centerdata/en.

[5] The RAND American Life Panel is modelled after the CentERpanel in the Netherlands.

voluntary deductible in that year. The survey was taken in October 2008 and the response rate was 68% (1,238 individuals). Table 3.A.2 compares survey means to the Dutch population based on data provided by Statistics Netherlands (CBS). The fraction of males among our panelists is high (69%), significantly higher than in the overall population (49%, Table 3.A.2). Apparently, males take household financial decisions more often than females do.

Figure 3.3 presents the sample distribution of voluntary deductible choice. For comparison reasons, we have also included the population distribution. The sample fraction of people choosing the lowest voluntary deductible is high (84%), yet lower than in the population (95%). This difference in choice patterns between the sample and the population is significant at the 99% confidence level. At least part of the explanation of this difference is that our sample consists of financial decision-makers, who by definition have greater interest in financial issues such as deductible choice in health insurance. Note that non-response is another potential explanation for the differences between the sample and population distribution of deductible choice.[6] We go further into the issue of non-response below.

Figure 3.3. Distribution of voluntary deductible choice in 2008, population and survey data



Note: Population data are from Vektis.

The individual level deductible choice data from our October 2008 survey were merged with existing DNB Household Survey (DHS) data, covering the same individuals. The DHS data include the discussed risk and risk-preference proxies, as well as several other personal characteristics that are used as control variables in the estimations. These other characteristics include number of children, whether the respondent has a partner (yes=1), living area (major urban=1) and highest education (1-6 scale, 6=university). After dropping observations with missing values for one or more of the explanatory variables, our data set consists of 947

[6] Non-response cannot fully explain the differences. If all non-respondents would in fact have the lowest voluntary deductible (€0), the sample share of zero voluntary deductibles would have been 88%, which is still lower than the population share of 95%.

observations for 2008. Though selection leads to a reduction in the sample size of 291 observations (24%), Figure 3 shows that the distribution of deductible choice is not markedly affected.

Table 3.3 gives sample averages of the explanatory variables, both for the group of respondents and for the total sample of 1,826 individuals (thus including non-respondents). Columns I, II, IV and V present sample averages after selection, i.e. dropping observations with missing values for one or more of the explanatory variables. Columns VI and VII provide sample average before selection. As a result of missing values, sample averages for several important regressors are not shown in Columns VI and VII.

Table 3.3 Sample means for different samples, before and after selection

	I	II	III	IV	V	VI	VII
	After selection					Before selection	
	Respondents				All	Respondents	All
	€0	> €0	t-test	All			
Age (years)	54.53	52.46		54.19	53.08	53.09	50.76
Gender (1=male)	0.70	0.83	**	0.72	0.70	0.70	0.69
Young men (1=male & age<45 yrs)	0.16	0.24	*	0.17	0.19	0.18	0.24
Health status (1-5, 5=poor)	2.17	2.07	*	2.16	2.17		
GP visits (number)	2.24	1.63	**	2.14	2.11		
Risk tolerance (1-7, 7= v. tolerant)	2.40	3.18	**	2.53	2.54		
Smoking (1=smoker)	0.16	0.25	**	0.16	0.18		
Drinking (1=drinker)	0.07	0.07		0.07	0.07		
Self-employed (1=yes)	0.07	0.13	**	0.05	0.08		
Portfolio share of equities	0.03	0.04		0.03	0.02		
Financial assets (€10,000)	3.68	6.20	**	4.09	3.76		
Annual income (€10,000)	4.11	5.02	**	4.26	4.18	4.19	4.13
Partner (1=yes)	0.73	0.70		0.72	0.70	0.72	0.72
Children (number)	0.60	0.49		0.58	0.63	0.66	0.75
Region (1=major urban)	0.16	0.23	*	0.17	0.17		
Education (1-6, 6=university)	3.82	4.06	*	3.86	3.87	3.86	3.89
Number of observations	*792*	*155*		*947*	*1,190*	*1,238*	*1,826*

Notes: Column III summarizes the two-sided t-test results of a comparison of the sample means of respondents with a voluntary deductible of €0 (shown in Column I) and individuals with a voluntary deductible higher than €0 (shown in Column II). ** and * indicate that the null of equal sample means is rejected at the 99% and 95% confidence level, respectively. The deductible choice data reflect 2008; the explanatory variables reflect 2007.

Focusing first on the sample averages after selection, we observe significant differences between those who chose a voluntary deductible larger than €0 (Column II) and those that did not (Column I). Consumers with a nonzero voluntary deductible are younger (though not significantly so), are more likely to be (young) males, bring fewer visits to the GP and typically regard themselves healthier than those with the lowest voluntary deductible. Hence, adverse selection appears relevant. People with a positive voluntary deductible are not significantly older, though. This is somewhat peculiar, since health care expenses clearly increase when people get older. Comparison also shows that consumers with above average deductibles are significantly more risk-tolerant, not only towards financial risk but also towards health and job risk. Among those with a voluntary deductible, the proportion of daily smokers and self-employed is significantly higher. However, the proportion of daily drinkers does not vary by deductible choice, nor does the portfolio share of equities. Wealth and income appear quite important to deductible choice, even though the size of the voluntary deductibles is tiny compared to the size of these variables.

Columns IV and V of Table 3.3 show the sample means (after selection) of respondents and all surveyed panelists, respectively, where the latter includes non-respondents. There does not appear to be much divergence between the two groups. However, when we estimate a probit model of the willingness to respond to the deductible choice survey, several regressors are statistically significant, including age, self-assessed health status and financial wealth (see Table 3.A.3 in Appendix 3.A). Such systematic non-response is a potential source of bias in the estimation of Equation (3.1). To assess the impact of non-response we will also estimate sample-selection models using full information maximum likelihood.

3.5 RESULTS AND DISCUSSION

3.5.1 Results

The results of (ordered) probit estimations of Equation (3.1) are presented in Table 3.4. The first four columns show the probit results; the last four columns give the ordered probit results. As there are only minor differences between the probit and ordered probit results, for example in terms of statistical significance, we discuss them together.

Starting with the risk variables, we find that these variables mostly have the expected signs, yet are not statistically significant or only marginally significant. In the probit specifications only gender seems a relevant risk driver. The null hypothesis that all risk proxies are irrelevant to deductible choice cannot be rejected at conventional confidence levels for the probit specifications.[7] In the ordered probit estimations, the number of GP visits is also statistically significant. Columns III and VII show that the young men dummy is insignificant in both the probit and the order probit specification. This result is hard to reconcile with expected-utility theory, since the expected out-of-pocket expenditures of young men are clearly below average. It is, however, consistent with the population data described in Section 3.2. In the population, the percentage of young men with a nonzero voluntary deductible is just slightly higher than the corresponding fraction in the overall population (8% versus 5%).

[7] A Wald test that the coefficients of the risk type proxies are all zero cannot be rejected at the 10% level.

Table 3.4. Deductible choice regressions

	I	II	III	IV	V	VI	VII	VIII
	Probit				Ordered probit			
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<u>Risk</u>								
Age	-0.00	(0.00)			-0.01	(0.00)		
Gender	0.27*	(0.14)			0.26*	(0.14)		
Young men (<45 years)			0.19	(0.14)			0.16	(0.14)
Self-assessed health status	-0.03	(0.09)	-0.00	(0.08)	-0.05	(0.08)	-0.03	(0.08)
SAHS*Age[a]	0.01	(0.01)			0.01	(0.01)		
GP visits	-0.03	(0.03)	-0.04	(0.03)	-0.05*	(0.03)	-0.06**	(0.03)
<u>Risk preferences</u>								
Financial risk tolerance	0.14***	(0.03)	0.14***	(0.03)	0.14***	(0.03)	0.15***	(0.03)
Financial assets + annual income	0.08***	(0.03)	0.09***	(0.02)	0.07***	(0.02)	0.08***	(0.02)
Smoking	0.31**	(0.13)	0.31**	(0.13)	0.26**	(0.13)	0.27**	(0.13)
Drinking	-0.18	(0.20)	-0.16	(0.20)	-0.19	(0.20)	-0.18	(0.19)
Self-employed	0.30*	(0.17)	0.32*	(0.17)	0.39**	(0.16)	0.40**	(0.16)
Portfolio share of equities	-0.35	(0.44)	-0.40	(0.44)	-0.35	(0.42)	-0.41	(0.41)
<u>Background</u>								
Partner	-0.17	(0.13)	-0.10	(0.13)	-0.19	(0.12)	-0.14	(0.12)
Number of children	-0.04	(0.06)	-0.03	(0.06)	-0.06	(0.06)	-0.05	(0.05)
Major urban	0.19	(0.13)	0.20	(0.13)	0.15	(0.13)	0.17	(0.13)
Education	-0.01	(0.04)	-0.01	(0.04)	-0.01	(0.04)	-0.01	(0.04)
Constant	-1.48	(0.35)	-1.65	(0.28)				
Log likelihood	-387		-389		-567		-570	
No. Obs	947		947		947		947	
Pseudo-R^2	0.08		0.08		0.07		0.06	

[a] Both variables are in deviation from their respective sample means. *Notes:* In the probit specifications, the dependent is 1 for individuals with a voluntary deductible higher than €0, and 0 otherwise. In the ordered probit specifications the dependent is voluntary deductible choice, which ranges from to €0 to €500. Estimated cut points of the ordered probit specification are not shown. Standard errors are in parentheses. ***, ** and * indicate significantly different from zero at the 99%, 95% and 90% confidence level, respectively.

Turning to the risk-preference variables, we observe that on the whole these variables have the expected sign (positive) and are highly significant. The null hypothesis that all risk preference proxies are statistically irrelevant to deductible choice is clearly rejected for all specifications.[8] Financial risk tolerance and wealth are both significant at the 99% confidence level. Interestingly, job risk and smoking behavior are also found to be significant determinants of deductible choice, even though self-employed individuals face greater background risk and despite increasing public awareness of the negative health effects and associated costs of smoking. The importance of our risk preference proxies in the regression results is especially noteworthy since these proxies have also been found to be relevant in decision-making when the

[8] The null that the coefficients of the risk preference proxies are zero is rejected at the 99% confidence level for all four specifications.

stakes are much larger (e.g., Barsky et al*.* 1997, Guiso and Paiella 2008). We will return to this point below. Note that in all specifications the background variables are insignificant.

The above findings are confirmed when we estimate Equation (3.1) as part of a Heckman-type sample selection model using full information maximum likelihood. Indeed, the sample selection model results, which are given in Table 3.A.4 in Appendix 3.A, closely resemble the above findings.

To gauge the economic importance of risk preferences and risk type in decision-making about deductibles, Table 3.5 gives predicted probabilities for the average individual, for relative low risk individuals, and for relatively risk-tolerant individuals. These predicted probabilities are based on the estimates in Table 3.4 (Columns I and V). The average individual and the low risk type (risk-tolerant type) differ only with respect to the stated risk type proxies (risk tolerance proxies). Compared to the average individual, a young man of 30 years old has – ceteris paribus - a slightly higher probability of choosing a nonzero voluntary deductible (18.3% versus 14.5%). The probability that a self-employed, smoking individual chooses a nonzero voluntary deductible is estimated at about 30%, which is more than twice that of the average adult. This probability increases further with financial risk tolerance and wealth, to above 50%. Hence risk preferences have a significantly greater impact on deductible choice than risk type, both statistically and quantitatively.

Table 3.5. Predicted probabilities of choosing a nonzero voluntary deductible

	Probit	Ordered probit				
	D>€0	*D*=€100	*D*=€200	*D*=€300	*D*=€400	*D*=€500
Average individual (covariates at mean)	14.5%	9.1%	2.3%	1.1%	0.0%	1.7%
Below average risk						
Young man, 30 years	18.3%	11.5%	3.1%	1.5%	0.5%	2.8%
with SAHS very good (p10)	19.2%	11.9%	3.3%	1.6%	0.5%	3.0%
with SAHS very good and zero GP visits (p10)	21.3%	13.3%	3.8%	1.9%	0.6%	3.8%
Above average risk tolerance						
Smoking, self-employed individual	29.9%	16.2%	5.1%	2.7%	0.9%	6.1%
with high financial risk tolerance (p90)	42.3%	19.8%	7.2%	4.1%	1.5%	11.4%
with high wealth and risk tolerance (p90)	50.6%	21.1%	8.3%	5.0%	1.8%	15.8%

Notes: The probabilities are predicted using the probit and ordered probit regression results given in Columns I and V of Table 3.4, respectively.

Since premium rebates may differ between health insurers, we use information on the actual rebates offered by health insurers to verify the robustness of our findings. Specifically, we construct a dummy variable which is 1 if the rebate offered for a voluntary deductible of €500 is higher than the expected costs, and 0 otherwise. The expected costs of a voluntary deductible of €500 are as shown in Figure 3.2. As not all panelists specified their basic health insurer, we lose 136 observations. Table 3.6 shows the results of this robustness exercise. For brevity's sake, only the estimated coefficients for the risk and risk-preference variables are shown. The newly constructed dummy has the right sign, yet is statistically insignificant. The risk-preference variables keep their significance and thus this robustness exercise further underpins our results.

Table 3.6. Deductible choice regressions, robustness exercise

	Probit		Ordered probit	
	Coeff.	s.e.	Coeff.	s.e.
Risk				
Rebate > expected costs	0.07	(0.20)	0.18	(0.19)
Self-assessed health status (SAHS)	-0.01	(0.09)	-0.03	(0.08)
GP visits	-0.04	(0.03)	-0.06**	(0.03)
Risk preferences				
Financial risk tolerance	0.15***	(0.04)	0.15***	(0.03)
Financial assets + annual income	0.08***	(0.03)	0.08***	(0.02)
Smoking	0.33**	(0.14)	0.30**	(0.13)
Drinking	-0.17	(0.21)	-0.21	(0.21)
Self-employed	0.37*	(0.19)	0.44**	(0.18)
Portfolio share of equities	-0.35	(0.44)	-0.35	(0.42)
Log likelihood	-329		-484	
No. Obs	811		811	
Pseudo-R^2	0.08		0.06	

Notes: See notes to Table 3.4.

3.5.2 Discussion

The above findings contrast with the standard expected-utility-of-wealth model in two respects. First, we find, at the most, modest evidence of adverse selection, while in the classical model, risk type is the only driver of deductible choice. Exemplarily, in this respect, is the deductible choice behavior of young men. From the population data we know that for the average young man it is clearly beneficial in expected value terms to choose a high voluntary deductible. In practice, however, very few young men choose to do so, both in the population and in our sample. Indeed, we find that, statistically, young men do not have a significantly higher probability of choosing a nonzero voluntary deductible. Our regression analyses suggest that this is caused by risk aversion. This brings us to our second contrasting finding, namely that risk preferences are a key determinant of deductible choice. This finding contrasts with the classical theory's prediction of local risk neutrality.

How then can we explain deductible choice behavior in the current context? Sydnor (2010) provides a number of potential explanations of why people tend to insure modest risks. Among them are risk misperception, consumption commitments and reference-dependent preferences.[9] Risk misperception may (partly) explain why people in general and young men in particular are reluctant to choose high deductibles. Indeed, anecdotal evidence suggests that, on the whole, people are not fully knowledgeable about which healthcare services are subject to out-of-pocket expenses, and which are not. What further complicates decision-making is that it is not reasonable to assume that out-of-pocket healthcare expenses are dichotomously distributed, i.e. individuals have either no costs or costs that exceed the deductible under consideration. The distribution of out-of-pocket expenses clearly shows this (see once more Table 3.A.1 in the appendix). By contrast, in studies of home insurance policies (Sydnor 2010) and auto insurance contracts (Cohen and Einav 2007), this assumption is explicitly made and indeed seems reasonable. As this simplification is not feasible here, risk misperception is presumably greater in health insurance. Nonetheless, misperception about the level of risk cannot explain the significance of risk preferences involved in deductible choice decision-making, as found in the regression analyses of Section 3.5.1. Indeed, even with risk misperception the classical model predicts that people are risk-neutral over small stakes.

Chetty and Szeidl (2007) argue that individuals can be significantly averse to moderate risks in the expected-utility model, namely when consumption commitments are taken into account. People tend to have consumption commitments, such as housing and durable goods, which are costly to adjust when adverse shocks occur. The authors show that these commitments raise the local curvature of the utility function, generating risk aversion to moderate risks. However, as Chetty and Szeidl's (2007) calibrations show, even with consumption commitments individuals are expected to be approximately risk-neutral to stakes in the order of $500. Accordingly, the existence of consumption commitments cannot satisfactorily explain the low demand for voluntary deductibles in the Netherlands. Corrected for average premium rebates, stakes in 2008 typically ranged from €150 (mandatory deductible, approximately $200) to below €450 (approximately $600).

While classical expected-utility models, either with or without consumption commitments, fail to fully rationalize risk aversion over small stakes, so-called reference-dependent utility models actually predict such preferences (Kőszegi and Rabin 2006, 2007). In reference-

[9] Sydnor (2010) discusses three other potential explanations of deductible choice behavior: borrowing constraints, role of sales agents and menu effects. With respect to borrowing constraints, Sydnor argues that it is not completely obvious how borrowing constraints would affect deductible choice. Indeed, extreme liquidity constraints may give an incentive to choose a high deductible, namely to save money up front, yet they may also give an incentive to choose a low deductible, as a liquidity-constrained individual would immediately run into trouble if an adverse scenario occurs. The remaining two explanations, role of sales agents and menu effects, do not seem very plausible in the current context. Most consumers (about 60%) contract their insurer via their employer, or another organization that negotiates certain collective benefits (ranging from sport clubs to internet groups), ruling out an influential role for sales agents. Menu effects would lead consumers to avoid extreme options, yet actual choice patterns show the opposite: Dutch adults typically choose the lowest voluntary deductible offered.

dependent models, which build from prospect theory (Kahneman and Tversky 1979, 1992), risky prospects are evaluated in isolation. Such decision-making has been labeled narrow framing, narrow bracketing, or myopia (Rabin and Thaler 2001). If small stake gambles are indeed evaluated in isolation and around a specific reference point, decision-making is dominated by gain-loss utility instead of the classical notion of outcome-based utility. The reference point is typically the status quo, which is being insured and having a €0 voluntary deductible in the current context, given the mandatory nature of basic health insurance and the low demand for nonzero voluntary deductibles. With this reference point, the payment of health insurance premium is planned and therefore not evaluated as a loss. Out-of-pocket expenses are, however, evaluated as losses. This brings us to a final important ingredient of reference-dependent utility models: loss aversion, i.e. agents are more sensitive to losses than they are to equivalent gains (e.g. Diecidue and Wakker 2001). In such a set-up, people are expected to be significantly risk-averse over modest stakes.

An appealing aspect of the Kőszegi and Rabin (2007) model is that it simultaneously allows for risk aversion over small and large stakes. This is because a person's utility is assumed to be the sum of outcome-based utility and gain-loss utility. Indeed, with wealth level w and reference wealth level r, a person's reference-dependent utility $u(w|r)$ is

$$u(w|r) = m(w) + \mu[m(w) - m(r)],$$

where $m(\cdot)$ is classical outcome-based utility, and $\mu(\cdot)$ represents gain-loss utility. Since outcome-based utility is approximately linear over small stakes, gain-loss utility dominates decision-making in the small. Over material stakes, however, decision-making is determined by the outcome-based part of a person's utility. Consequently, over large stakes, risk aversion is driven by the traditional mechanism of diminishing marginal utility of wealth, while over small stakes, risk aversion is the result of loss aversion.

Combining our results with existing empirical work, it appears that risk attitudes over small and large stakes are closely related, or stated differently, manifestations of the same preferences in different domains. This is because the risk-preference proxies we use to explain deductible choice have been found, by others, to be relevant determinants of choice behavior over much larger stakes. For example, Guiso and Paiella (2006) find that lifetime wealth (i.e. the sum of financial wealth and income) and job risk are positively related to risk tolerance over stakes in the order of €5000 (approximately a factor 10 of the stakes in deductible choice). Bertaut (1998) and Alessie, Hochguertal, and Van Soest (2004) show that equity ownership increases with wealth. Cutler et al. (2008) find that smokers are less likely to buy acute health insurance, leaving them more exposed to substantial financial risk. Hence, it seems that individuals that are more risk-tolerant to large stakes are also more risk-tolerant to small stakes, and vice versa. Establishing the importance of the relationship between risk taking in the small and in the large is an interesting topic for future research.

3.6 CONCLUSION

In this chapter we study the incentives behind deductible choice in Dutch universal health insurance. The unique institutional characteristics make the Dutch health insurance market close to ideal for studying small stakes risk taking over in a real-life setting. Health insurers in the Netherlands are obliged to accept residents at community-rated premiums, and Dutch residents are obliged to buy the basic health insurance policy. Consequently, risk selection by health insurers is impossible and consumer choice is restricted to choosing a health insurer and a deductible-rebate package with that insurer.

According to standard expected-utility theory, people are roughly risk-neutral over small stakes. Since premiums are not risk-based, risk-neutral preferences imply that healthier individuals have more incentives to choose a nonzero voluntary deductible, leading to adverse selection. Using population and survey data, we show that this is indeed the case, yet healthy individuals' monetary incentive to choose a high voluntary deductible is more often than not dominated by their risk aversion. Exemplarily is the choice behavior of young men (aged 44 years and under). While young men have low out-of-pocket health expenses, and would therefore benefit in expected-value terms from a high deductible, 92% of them choose the lowest deductible possible. Regression analyses strongly suggest this choice behavior is caused by risk aversion. Indeed, corrected for risk type, we find that more risk-averse individuals are significantly less likely to opt for a high deductible, and vice versa.

Risk aversion over small stakes can be rationalized by reference-dependent utility models, where narrow framing and loss aversion play pivotal roles. The results of this chapter indicate there is value in exploring such utility models further. That people are risk averse to minor risks explains the existence of insurance market for such risks. For example, it is not unusual for buyers of cellular phones or tablet computers to insure their new possession.

APPENDIX 3.A TABLES

Table 3.A.1. Out-of-pocket health expenditures with voluntary deductible €0 in 2008

			Out-of-pocket expenditures (€)									
Age (years)	Gender	# of people	Avg.	p10	p20	p30	p40	p50	p60	p70	p80	p90
20-24	male	443,892	62	0	0	0	9	21	61	146	150	150
25-29	male	448,516	61	0	0	0	8	19	57	145	150	150
30-34	male	448,767	65	0	0	0	11	29	79	150	150	150
35-39	male	549,672	70	0	0	0	19	43	108	150	150	150
40-44	male	569,307	77	0	0	9	23	64	149	150	150	150
45-49	male	549,289	85	0	0	15	41	108	150	150	150	150
50-54	male	499,984	96	0	0	29	93	150	150	150	150	150
55-59	male	482,804	108	0	19	81	150	150	150	150	150	150
60-64	male	462,043	118	0	52	150	150	150	150	150	150	150
65-69	male	337,467	127	18	131	150	150	150	150	150	150	150
70-74	male	269,236	135	67	150	150	150	150	150	150	150	150
75-79	male	204,183	139	145	150	150	150	150	150	150	150	150
80-84	male	127,210	141	150	150	150	150	150	150	150	150	150
85-89	male	61,991	140	150	150	150	150	150	150	150	150	150
90-94	male	20,288	138	123	150	150	150	150	150	150	150	150
20-24	female	441,833	102	14	36	57	88	148	150	150	150	150
25-29	female	453,296	106	12	38	63	109	150	150	150	150	150
30-34	female	461,038	108	9	38	69	135	150	150	150	150	150
35-39	female	565,176	104	0	30	60	113	150	150	150	150	150
40-44	female	575,383	103	0	26	56	110	150	150	150	150	150
45-49	female	564,953	107	0	29	70	143	150	150	150	150	150
50-54	female	519,402	113	0	38	106	150	150	150	150	150	150
55-59	female	493,903	118	0	54	147	150	150	150	150	150	150
60-64	female	473,590	123	12	94	150	150	150	150	150	150	150
65-69	female	355,351	131	34	150	150	150	150	150	150	150	150
70-74	female	308,546	137	94	150	150	150	150	150	150	150	150
75-79	female	272,499	140	150	150	150	150	150	150	150	150	150
80-84	female	215,405	141	150	150	150	150	150	150	150	150	150
85-89	female	139,512	138	147	150	150	150	150	150	150	150	150
90-94	female	71,071	133	35	150	150	150	150	150	150	150	150

Notes: Data are from Vektis. Column 'Avg.' shows the average expenditures; columns denoted "p10", "p20", etc. show the respective percentiles of the expenditure distributions. Total out-of-pocket expenditures of people with a €0 voluntary deductible are bound by the mandatory deductible of €150. Note that similar data are available for the other voluntary deductible categories (€100, €200, …, €500).

Table 3.A.2. Comparison of survey and population means, 2008

Variable	Survey mean	Population mean
Age (years)	***53.09	49.06
Gender (1=male)	***0.70	0.49
Education (1-6, 6 = university)	3.86	3.80

Source for population means: Statistics Netherlands (CBS) data .
Notes: Survey means are for respondents before selection (1,238 observations). For variables age and gender, we present population means for people of 20 years and over. For education, the CBS-data cover 15-65 year olds. *** indicate significantly different from zero at the 99 confidence level between survey and population means.

Table 3.A.3. Probit for the probability of answering the deductible choice question

Variable	Coefficient	s.e.
Age	0.01***	(0.00)
Gender	0.01	(0.11)
Self-assessed health status (SAHS)	-0.13**	(0.07)
SAHS*Age[a]	-0.01	(0.00)
GP visits	0.02	(0.02)
Financial risk tolerance	-0.02	(0.03)
Financial assets + annual income	0.06**	(0.03)
Smoking	0.07	(0.11)
Drinking	-0.02	(0.17)
Self-employed	-0.00	(0.15)
Portfolio share of equities	0.89*	(0.54)
Partner	0.26**	(0.10)
Number of children	-0.03	(0.04)
Major urban	-0.05	(0.11)
Education	-0.03	(0.03)
Constant	0.22	(0.28)
Log likelihood	-575	
No. Obs	1190	
Pseudo-R^2	0.05	

[a] Both variables in this interaction term are in deviation from their sample means. *Notes*: The left-hand-side variable is an indicator that is equal to 1 if the household financial decision-maker has responded to the question on deductible choice. ***, ** and * indicate significantly different from zero at the 99%, 95% and 90% confidence level, respectively.

Table 3.A.4. Deductible choice regressions, selection models

	I	II	III	IV
	Probit		Ordered probit	
	Coeff.	s.e.	Coeff.	s.e.
Risk				
Age	-0.00	(0.00)	-0.00	(0.00)
Gender	0.20	(0.12)	0.18	(0.12)
Self-assessed health status (SAHS)	-0.02	(0.08)	-0.04	(0.03)
GP visits	-0.03	(0.02)	-0.04	(0.03)
Risk preferences				
Financial risk tolerance	0.12***	(0.03)	0.12***	(0.04)
Financial assets + annual income	0.08***	(0.03)	0.08***	(0.02)
Smoking	0.33***	(0.12)	0.30**	(0.12)
Drinking	-0.15	(0.19)	-0.16	(0.19)
Self-employed	0.27*	(0.16)	0.35*	(0.18)
Portfolio share of equities	-0.22	(0.43)	-0.23	(0.41)
ρ	0.97	(1.28)	0.81	(1.64)
No. obs	1190		1190	
Censored obs.	243		243	
Pseudo-R^2	0.07		0.05	

Notes: See notes to Table 3.4 in the text. Table 3.A.3 specifies the selection equation. Variables 'partner', 'number of children', 'major urban' and 'education' were excluded from the outcome equation to lessen multicollinearity. ρ is the correlation between the disturbances of the outcome and selection equation.

Chapter 4

RESTRUCTURING OF THE DUTCH NON-LIFE INSURANCE INDUSTRY[*]

[*] This chapter is based on Bikker and Gorter (2011).

4.1 INTRODUCTION

The European Union's Single Market Program (SMP) has had a profound impact on the financial services landscape in Europe, and in particular on Europe's insurance markets. Before the SMP, national insurance markets in Europe were essentially closed to cross-border competition. In many jurisdictions, the scope for product and price competition between insurance firms was also limited. Through a series of directives, and in particular the implementation of the Third Generation Insurance Directives (TGID) in 1994, the European insurance industry was deregulated (with the exception of solvency regulation). As a result, European insurance firms are now allowed to operate across national boundaries via a so-called European Single License, and free to develop new insurance products and set prices at their discretion. The principal goal of deregulation in general and the EU's SMP in particular is to improve market efficiency and enhance consumer choice through increased competition.[1]

In formerly highly-regulated European countries, insurance deregulation meant a distinct break with the past. A prime example is Germany where prices used to be regulated for the entire industry, enabling the most inefficient providers to remain in the market.[2] By contrast, in traditionally more liberal countries, like the Netherlands and the UK, the regime change had limited direct impact. The TGID transmitted a regulatory model similar to that of the Netherlands across Europe, fostering a level playing field, except where solvency regulation was concerned.

Even so, the structure of the Dutch non-life insurance industry has changed considerably in recent years. Between 1995 and 2005, the number of non-life insurers registered with the Dutch supervisor dropped by more than 20%, and average premium income almost doubled in real terms. Interestingly, this market consolidation did not lead to widespread conglomeration of the industry. On the contrary: the market share of focused Dutch non-life insurers – monolines active in one line of business only – actually increased during these years. At the same time, though relatively few demutualizations (i.e. mutual firms converting to stock charter) occurred, the market share of mutual insurers dropped substantially.

This chapter investigates the restructuring of the Dutch non-life insurance industry from a cost-efficiency point of view. Given the objective of the TGID to increase the efficiency of Europe's insurance markets, we expect efficiency considerations played a central role. As

[1] The deregulation of the European insurance industry is discussed in Swiss Re (1996).

[2] In fact, efficient German insurance firms used to face a penalty for good performance. Before 1994, profits over 3% of premium income had to be divided between policyholders and shareholders, with at least 90% going to policyholders through terminal bonuses (Rees and Kessner 1998). Although this regulatory scheme created a stable and transparent market for consumers, price competition was weak, and product choice was limited.

mentioned, one of the key trends in the Dutch market has been industry consolidation; what cost incentives could have been driving this development? An obvious motive for consolidation is the existence of economies of scale. When there are increasing returns to scale, firms have an incentive to engage in M&A activities. Another potential rationale for industry consolidation derives from X-efficiency differences between insurance firms. X-efficiency reflects firms' ability to drive down production costs, controlled for output volumes and input price levels (Leibenstein 1966). Insurers that are relatively X-inefficient are a natural purchase target for insurers with more skilled management. Consequently, we expect average X-efficiency has improved over the 1995-2005, also because the threat of a hostile takeover encourages managers to increase their effort and take up internal slack.

Besides consolidation, more strategic focus has also been an important development in the Dutch non-life insurance industry. Strategic focus can either increase or decrease efficiency. The strategic-focus hypothesis states that focus on types of insurance activities adds value, through reduction of owner-manager incentive problems (Jensen 1986, Meyer, Milgrom, and Roberts 1992). The competing conglomeration hypothesis argues conversely, i.e. conglomeration – as opposed to focus – creates firm value from exploiting economies of scope (Teece 1980). The observed increase in focus suggests that in non-life insurance the benefits outweigh the costs, giving insurance firms an incentive to streamline their organizations.

The third key trend has been the declining importance of the mutual organizational form. The organizational form of insurers is hypothesized to affect cost efficiency mainly via comparative advantages in dealing with agency costs. While the stock ownership form is more appropriate in controlling owner-manager conflicts, mutual ownership helps mitigate conflicts between owners and policyholders. Besides differences in dealing with agency problems, stock-owned insurers have an additional advantage in their superior access to capital. The efficient-structure hypothesis predicts that mutual and stock insurers are sorted into market segments where their respective comparative advantages materialize (Mayers and Smith 1981). Conversely, the expense-preference hypothesis predicts that mutual insurers will generally be less cost-efficient than stock insurers, because the mutual ownership form provides weaker mechanisms for controlling owner-manager agency conflicts (Mester 1989).

Hypotheses about consolidation, strategic focus, and organizational form have been widely investigated in the insurance efficiency literature.[3] We contribute to this literature by providing

[3] Cummins and Weiss (2000) provide a comprehensive review of the literature. A more recent literature overview is given by Eling and Lunen (2010).

evidence on these hypotheses for the Dutch non-life insurance industry, currently the seventh market worldwide in terms of premium income (Swiss Re 2012).[4]

We analyze the Dutch non-life insurance industry over the period 1995-2005 using a translog cost function (TCF). Our TCF includes several dummy variables to provide insight into the cost-efficiency effect of organizational form and focus. Because cost levels vary greatly across insurers, we employ the thick frontier approach (TFA) of Berger and Humphrey (1991) to reveal X-efficiencies and their development over time. We apply our methodology to a comprehensive and unique data set that covers all non-life insurers licensed by DNB.

The rest of this chapter is structured as follows. The next section reviews the hypotheses to be tested and provides theoretical background to the empirical work. The third section describes the structure of the Dutch non-life insurance industry as well as our data and methodology. The fourth section presents the empirical results. The last section concludes.

4.2 HYPOTHESES

4.2.1 Consolidation: scale economies and X-efficiency

Since the midst of the 1990s, numerous M&As in the Dutch non-life insurance industry have led to a substantial increase in average firm size. In justifying M&As, insurers often cite the presence of scale economies. Scale economies (or increasing returns to scale) are present if the unit costs of production decline as firm size increases, and derive from the spreading of fixed costs over a broader output base. In the provision of insurance services, relatively fixed production factors are ICT hardware and software, managerial expertise and financial capital (Cummins and Rubio-Misas 2006). In light of the observed consolidation, we hypothesize the following.

H4.1: Non-life insurance operations in the Netherlands are generally characterized by increasing returns to scale (scale economies).

Under increasing returns to scale, larger firms are more scale-efficient than smaller firms. Given that the average insurer has sized up its scale of operations and presuming that scale economies decrease in size, we expect that scale efficiency in the Dutch non-life insurance industry has improved.

[4] Before 2006, the Dutch non-life insurance sector ranked ninth worldwide in terms of premium income (Swiss Re 2006). The large increase in premium volume since 2006 is the direct result of the Health Insurance Act that came into effect 1 January 2006 and privatized the former national health insurance funds.

H4.2: Unused scale economies have decreased over our sample period (increasing scale efficiency).

Consolidation also has the potential of increasing the X-efficiency of an industry. Presuming that managers of acquiring non-life insurers are generally more capable than those of targets, that inefficient firms generally withdraw from the market, and that the threat of a takeover increases when market consolidation picks up, we expect X-efficiency to have increased over time.

H4.3: X-efficiency of Dutch non-life insurers has increased (increasing X-efficiency).

4.2.2 Organizational form

The stock and mutual organizational form are present in most developed insurance markets around the world (Swiss Re 1999). In the Dutch insurance market, too, stock and mutual insurers have coexisted for many decades. Agency theory explains the coexistence of stock and mutual insurers in terms of their relative success in dealing with specific types of incentive conflicts (Mayers and Smith 1988). The stock ownership form has a comparative advantage in controlling owner-manager conflicts. Stock-owned firms have alienable ownership claims, facilitating managerial control mechanisms such as proxy fights, performance-related remuneration packages and hostile takeovers that help reduce inefficient behavior by managers. The mutual ownership form offers weaker control mechanisms. Mutual insurers are however more successful in managing owner-policyholder conflicts, as these are effectively internalized. Besides these comparative advantages in dealing with agency costs, the stock organizational form has an additional advantage: its superior access to capital. While mutual insurers depend on retained earnings as the primary source of new capital, stock insurers can also raise capital directly in the markets or receive capital infusions from publicly traded parents.

According to the efficient-structure hypothesis, stock and mutual insurers are successful in lines where they have comparative advantages. Mayers and Smith (1981) argue that the stock-ownership form is more appropriate in lines that require a large degree of managerial discretion, considering its advantage in controlling owner-manager conflicts. Such lines include commercial coverage, where pricing and underwriting typically are customized, and managerial discretion is often needed (Cummins, Rubio-Misas, and Zi 2004). On the other hand, the mutual ownership form is likely to be more successful in lines of business that require limited managerial discretion. In such lines, policies are typically standardized, loss data are extensively available, the variance of losses is comparatively small (Lamm-Tennant and Starks 1993, Doherty and Dionne 1993) and screening is less valuable (Hansmann 1985, Smith and Stutzer 1990). The preceding suggests the following hypothesis.

H4.4: *Stock and mutual insurers are relatively successful in lines of business where each has comparative advantages (efficient-structure hypothesis).*

Another organizational form hypothesis we investigate in this chapter is the so-called expense-preference hypothesis (Mester 1989). This hypothesis predicts that mutual insurers are less efficient than stock insurers. The reasoning is that as utility-maximizing managers have a preference to spend more on staff, office furniture and other perquisites (Williamson 1963), mechanisms are needed to control managerial opportunism. And, as mentioned above, the stock-ownership form offers superior mechanisms for controlling inefficient managerial behavior. Whether expense-preference behavior exists or not is ultimately an empirical question. Our fifth hypothesis states as follows.

H4.5: *Mutual insurers are less cost-efficient than stock insurers (expense-preference hypothesis).*

4.2.3 Strategic focus

Restructuring has brought about more focus within the Dutch non-life insurance industry, where focus is measured by firms' line-of-business concentration. Theory is unclear about the efficiency effect of increased focus. On the one hand, it is argued that focus adds value by reducing agency problems, where underperforming activities are cross-subsidized and negative net present value investments are made (Jensen 1986, Meyer, Milgrom, and Roberts 1992). On the other hand, it is also argued that focus reduces firm value, because cost and/or revenue scope economies are not exploited. Cost scope economies arise from the sharing of production inputs across multiple activities (Teece 1980). Revenue scope economies may be realized through providing 'one-stop shopping' to consumers who are willing to pay more for the added convenience.

So, theoretically there are both benefits and costs associated with organizational focus. While proponents of the strategic-focus hypothesis argue that the benefits outweigh the costs, proponents of the competing conglomeration hypothesis argue conversely (Berger, Cummins, Weiss, and Zi 2000). In view of the rise of monolines over our sample period, we expect a positive relationship between focus and cost-efficiency, consistent with the strategic-focus hypothesis.

H4.6: On average, focused non-life insurers are more cost-efficient than diversified ones (strategic-focus hypothesis).

4.3. INDUSTRY STRUCTURE, DATA AND METHODOLOGY

This section discusses the structure of the Dutch non-life insurance industry and how it has developed over time. Here, we also describe the data as well as our modeling and estimation methodology.

4.3.1 Industry structure

The Netherlands has quite a sizeable non-life insurance industry, especially since the privatization of the former health insurance funds in 2006. But even before that, the Dutch non-life market ranked among the top ten non-life markets in the world (Swiss Re 2006). We analyze the Dutch non-life industry over the 1995-2005 period, which was characterized by considerable change. During these years, the number of non-life insurers registered with the Dutch supervisor dropped by more than 20%. The number of foreign-licensed non-life insurers increased, however. While large in number (540 in 2003), foreign-licensed firms are estimated to account for less than 2% of non-life business in the Netherlands (Oosenbrug 2007). This being so, the mere entry of a large number of foreign-licensed firms has presumably stimulated competition among the incumbent firms.[5] We do not consider foreign-licensed firms in our analysis since data about these firms are lacking.

Intermediaries are the most important distribution channel of non-life insurance products in the Netherlands, typically capturing more than 50% of annual premium income. The exception is accident and health insurance, where direct writing is more common (see Figure 1.3 in Chapter 1). This percentage is high compared to that in other large European markets such as Germany and the UK, where less than 20% is sold via intermediaries. Only a minor share of Dutch non-life policies (about 10%) is sold via related banks.

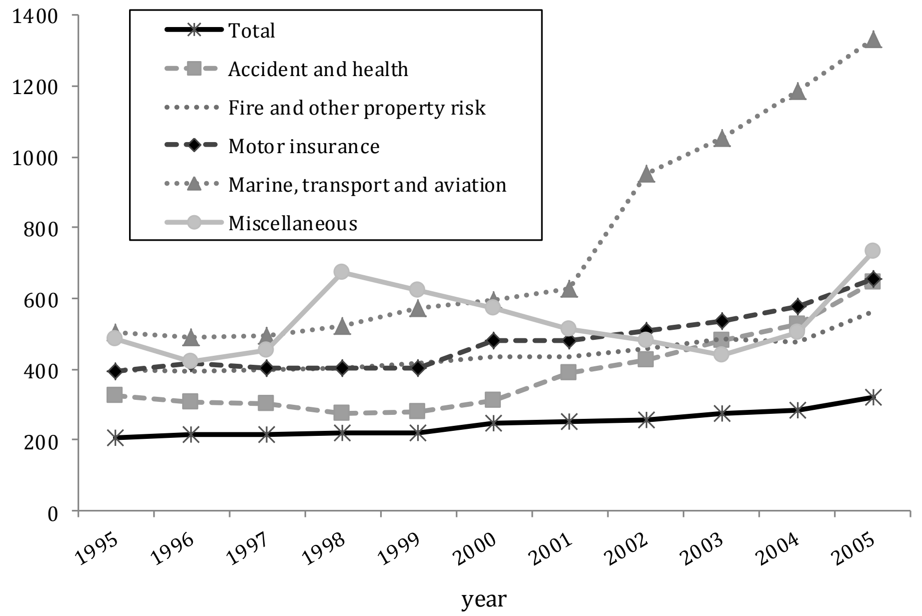
The non-life industry is composed of five lines of business: (i) accident and health, (ii) motor, (iii) marine, transport and aviation, (iv) fire and other property risk, and (v) miscellaneous insurance. Accident and health insurance is the most important line of business.

[5] Mahlberg and Url (2000) and Swiss Re (2000) argue that European deregulation has intensified competition primarily in domestic markets (via the establishment of subsidiaries) rather than through true cross-border competition. Differences in taxation and contract law, cultural heterogeneity, and informational advantages have been cited as reasons for weak cross-border competition (Cummins and Rubio-Misas 2006).

In 2005 it accounted for 48% of total premiums, whereas motor insurance and fire and other property risk insurance had shares of 20% and 14%, respectively, leaving 18% for the remaining lines of business.

As mentioned, on-going consolidation has been an important feature of the Dutch non-life industry. Figure 4.1 presents Herfindahl-Hirschman concentration indices (HHIs), defined as the sum of squared market shares (in percentages), based on premium income per line of business over 1995-2005, and calculated at the firm level. The indices increase substantially over the sample period, both for the entire industry and for the individual lines of business. In 2005, the marine, transport and aviation insurance line was most consolidated; in this line competition and efficiency are fostered by the presence of well-informed and cost-conscious buyers (businesses), some of which have the alternative of self-insuring.

Figure 4.1. HHIs Dutch non-life insurance industry (aggregated as well as by line)



Notes: HHIs are the Herfindahl-Hirschman concentration indices, defined as the sum of squared market shares in percentages, based on premium income.

The upper panel of Table 4.1 presents market shares of the largest 5, 10 and 20 firms for the total non-life insurance industry, based on net premium income. Each of these measures of market concentration increased between 1995 and 2005. The top five insurers account for virtually the entire rise in market concentration, indicating that they dominated the restructuring of the non-life insurance industry. Furthermore, we observe that the market share of firms that operate in two or more lines of business – i.e. multiline insurers – has slightly decreased over the sample period, while the market share of monoline insurers has slightly increased (columns six and seven of Table 4.1, upper panel). This shows that market consolidation did not lead to widespread conglomeration of the industry.

The last column of Table 4.1 (upper panel) displays the market share of mutual insurers, which has declined substantially over the 1995-2005 period, from 23% to 14%. The overall and increasing dominance of stock insurers supports the efficient-structure hypothesis (*H4.4*). Upon the introduction of the TGID, the Dutch non-life insurance sector entered a period of dynamism and growth. Product innovation became more important, requiring significant capital investments. In such market circumstances, extensive managerial discretion and ready capital access are called for. In both respects, stock-owned firms have an advantage over mutual firms.

Table 4.1. Non-life insurance groups and firms, and their respective market shares

Market share of groups and firms over time (in %)

Year	Groups	Largest 5 firms	Largest 10 firms	Largest 20 firms	Monoline firms	Multiline firms	Stock firms	Mutual firms
1995	78	24	41	58	38	62	77	23
2000	79	30	45	63	37	63	84	16
2005	76	33	51	71	40	60	86	14

Number of groups and firms over time

Year	Groups	Affiliated firms	Unaffiliated firms	Monoline firms	Multiline firms	Stock firms	Mutual firms
1995	27	109	68	103	74	107	70
2000	24	117	78	127	68	103	92
2005	18	90	68	108	50	82	76

Market share of stock and mutual firms by line of business (in %)

Org. form	Accident and health	Fire and other prop. risk	Motor	Marine, transport and aviation	Miscel-laneous
Stock	71	92	88	89	91
Mutual	29	8	12	11	9

Notes: 'Org. form' is an abbreviation for 'Organizational form'. Figures shown are after selection. See Subsection 'Data' for the selection criteria.

The middle panel of Table 4.1 shows the number of non-life insurance firms and groups. During the sample period, the number of non-life groups decreased from 27 to 18, while the market share of such groups declined only slightly (from 78% to 76%). Average group size thus increased substantially, as groups took over other groups and, hence, their affiliated firms. This is also reflected by the decline of the number of affiliated firms over our eleven year sample period. By contrast, the number of unaffiliated firms remained fairly stable over the years. Apparently, consolidation was strongest within insurance groups. We observe that monolines far outnumber multilines. The number of monoline insurers is relatively high, because they are both easier to establish (fewer regulatory requirements) and easier to manage. Considering developments over time, we observe that the number of multiline firms decreased by 32% in eleven years while the number of monolines increased by 5%. Again, we find that consolidation is stronger among multiline firms. Interestingly, while the number of stock firms fell, the number of mutual firms actually increased. So, given the development of the respective market shares of stock and mutual insurers, the former apparently scaled up much more than the latter.

The lower panel of Table 4.1 shows that all lines of business are dominated by stock firms. The fact that mutual insurers' most successful line of business is accident and health insurance (29% market share) is consistent with the efficient-structure hypothesis (*H4.4*). Accident and health insurance, and in particular health insurance, is a relatively stable, personal line of business with standard underwriting practices. Moreover, the variance of losses in health insurance are restricted on account of the Dutch 'Exceptional Medical Expenses Act', by which exceptionally high medical expenses are publicly insured. So, in accident and health insurance, compared to the rest of the non-life industry, a smaller amount of managerial discretion is warranted and with that there is less scope for owner-manager conflicts. Conversely, there is ample scope for owner-policyholder conflicts in this line of business. The mutual form has a comparative advantage in dealing with such agency conflicts.

4.3.2 Data: outputs, inputs and sample selection

Outputs. The output of insurers consists mainly of unobservable services. Insurance output must therefore be measured using proxy variables. In the literature, there is little agreement on the appropriate measure of non-life insurance output. Both premiums and losses incurred have been used several times, and both have their shortcomings. The main flaw of premiums is that they represent revenues, i.e. prices times quantities, rather than volumes (Yuengert 1993). Losses incurred do not suffer from this drawback. On top of this, losses incurred is a metric that is broadly consistent with theory, as losses capture the risk-pooling and risk-bearing services of non-life insurers. Nonetheless, this empirical proxy also has its drawbacks (see Cho 1988). First, the stochastic nature of non-life losses is likely to create an 'errors-in-variables problem', particularly where the insurance includes catastrophe risk. Second, when losses incurred are used, the output quality of loss control and risk management is undesirably ignored. So, given that both premiums and losses incurred are imperfect measures of insurance output, we follow the suggestion of Cummins and Weis (2000) and alternate between both measures.[6] To eliminate reinsurance activities from our analysis, we use premiums net of reinsurance ceded and losses net of reinsurance received. Furthermore, to proxy for the intermediation function of non-life insurers, we use total investments as an additional output variable.

Inputs and input prices. Following the recent insurance-efficiency literature, we take labor, equity capital and debt capital as inputs. The price of labor is measured by average labor costs per full time equivalent employee in the Dutch insurance sector, provided by Statistics Netherlands (CBS). Using a firm-specific labor costs variable would be preferable, yet the data needed are not available. Our other two inputs are financial capital inputs. Ideally, the cost of equity capital is determined from traded share prices. However, as few Dutch non-life insurers firms issue traded shares, we proxy the cost of equity capital with the total return on the Amsterdam Stock Exchange (AEX) Index. Specifically, we use the average rate of return on the AEX Index for the ten-year period foregoing each year of the sample period. Debt capital is an additional source of funds and consists of borrowings from creditors and deposits by reinsurance firms. To proxy the cost of debt capital, we use the one-year Dutch Treasury bill rate (cf. Cummins et al. 2004, Cummins and Rubio-Misas 2006). Hence, our input prices change over time, not across insurance firms.

Sample selection. Our database primarily consists of detailed regulatory statements of all non-life insurers licensed by DNB, central bank and prudential supervisor of the Netherlands.[7] The data cover the period 1995-2005. While the initial data set consists of 2994 firm-year observations, we have excluded 846 observations because of lacking data, on account of zero or negative output or cost variables, where costs exceeded premiums (e.g. run-off firms), or because equity capital was either implausibly low (less than 5% of total assets) or implausibly high (more than 95% of total assets).[8] We also excluded firms providing less than three consecutive years of available data. In monetary value terms, the importance of excluded firms is relatively small.

[6] Note that we use non-discounted net losses incurred as output measure. While the present value of real losses incurred would be a more appropriate output measure, we are faced with the empirical difficulty that non-life insurers payout proportions are unavailable in the Netherlands. For that reason we are unable to discount losses incurred, at least not in a reliable and consistent way. Other researchers on European insurance markets have faced this difficulty as well. Indeed, Cummins et al. (2004), Cummins and Rubio-Misas (2006), and Fenn, Vencappa, Diacon, Klumpes, and O'Brien (2008) also use non-discounted losses to measure non-life insurance output.

[7] Note that the supervisory system in the Netherlands changed in 2004. Before 2004, insurers were licensed by the 'Pensioen- & Verzekeringskamer' (Pensions and Insurance Chamber). After the integration of the Pensions and Insurance Chamber into the Dutch central bank, DNB issues licenses to insurers in the Netherlands.

[8] The data does not distinguish between missing data and zero values. We have excluded 34 observations with negative values.

Exclusion from the analysis concerns about 13% of the sum total of net premiums. The data set is an unbalanced panel due to mergers and acquisitions, terminations, new entrants and our selection procedure.

Table 4.2 (upper panel) gives summary statistics of our sample of Dutch non-life insurance firms. By all size measures, stock insurers are significantly larger than mutual insurers. The average stock firm has total assets of about 210 million euros (at 1995 prices), which is more than four times larger than that of the average mutual (47 million euros). The net premiums and losses of stock firms are also roughly four times those of mutual firms. The ratio of equity capital to total assets shows that stock-owned insurers are significantly more leveraged than mutual firms, probably reflecting stock insurers' greater flexibility in raising fresh capital when needed. The HHI variable in the final row of Table 4.2 is based on the shares of output associated with the different lines of business within each firm and thus measures business specialization. A lower value for this specialization measure reflects a more diversified product mix. Stock firms are significantly more diversified than mutual firms, which are predominantly active as monoline insurers.

Table 4.2. Sample averages at firm level (1995-2005)

Variables	Pooled	Stock	Mutual	t-test
Net premiums – output	63,431	98,834	24,571	**
Net losses – output	48,372	73,067	21,265	**
Investments – output	103,664	164,988	36,353	**
Equity capital – input	38,655	52,002	24,004	**
Debt capital – input	13,335	20,913	5,016	**
Price of labor	47	47	47	
Price of equity capital (%)	34.28	34.28	34.28	
Price of debt capital (%)	3.14	3.14	3.14	
Total costs	16,915	29,105	3,536	**
Total assets	132,287	209,823	47,178	**
Equity capital / Total assets	0.45	0.35	0.55	**
Debt capital / Total assets	0.14	0.11	0.16	**
HHI by line (based on premiums)	0.82	0.69	0.97	**
Number of firms	*195*	*102*	*93*	

** Indicates sample means of stock and mutual firms are significantly different at the 99% confidence level. *Note:* monetary values are expressed in thousands of euros, deflated to the 1995 price level using the Dutch consumer price index.

4.3.3 Methodology

To test the formulated hypotheses above, we use a translog cost function (TCF). Christensen, Jorgenson, and Lau (1973) proposed the TCF as a second-order Taylor expansion, usually around the mean, of a generic function with all price and output variables appearing as logarithms. The TCF is a flexible functional form that has proven to be an effective tool for the empirical assessment of efficiency. For a theoretical underpinning and an overview of applications in the literature, see Bikker and Van Leuvensteijn (2008). Our TCF reads as follows:[9]

$$\ln C = \alpha + \sum_{i=1}^{2} \beta_i \ln Y_i + \sum_{k=1}^{3} \gamma_k \ln P_k + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \beta_{ij} \ln Y_i \ln Y_j + \frac{1}{2}\sum_{k=1}^{3}\sum_{l=1}^{3} \gamma_{kl} \ln P_k \ln P_l$$
$$+ \sum_{i=1}^{2}\sum_{k=1}^{3} \delta_{ik} \ln Y_i \ln P_k + \kappa HHI + \rho STOCK + \sum_{q=1}^{4} \left[\phi_q + \psi_q STOCK\right] LOB_q + \lambda GROUP + \tau T$$
$$+ \varepsilon, \qquad (4.1)$$

where C are total costs, comprising both operating expenses and distribution or acquisition costs. In Equation (4.1) the two Y variables are the outputs, whereas the three P variables are the input prices. Variable HHI measures line-of-business specialization (defined as in Table 4.2) and is included to test the strategic-focus hypothesis. Under *H4.6* we expect a minus sign for parameter κ. Variable $STOCK$ is a dummy variable that takes the value 1 for stock firms and 0 otherwise, and captures cost differences between stock and mutual firms. The line-of-business (LOB) variables are dummies for four of the five lines of business (miscellaneous insurance is the reference group), which take the value 1 if the firm is predominantly active in the respective line of business. These dummies capture any remaining time-invariant (constant) effects related to the respective lines of business. We also include interaction terms between variables $STOCK$ and LOB to investigate cost differences between stock and mutual insurers over the lines of business. Dummy variable $GROUP$ (1 for firms affiliated to a group) is included to control for group affiliation, as the group structure may also affect the cost level. To account for a possible time trend in the insurance industry's cost level, as an approximation of technical progress, a time variable ($T = 0,1,2,\ldots,10$) is included in the TCF. The error term is represented by ε.

Equation (4.1) is estimated by ordinary least squares (OLS), where the standard restrictions of symmetry and homogeneity of degree one in input prices are imposed in estimation (i.e. $\beta_{1,2} = \beta_{2,1}$; $\gamma_{kl} = \gamma_{lk}$ for $k,l = 1,2,3$; $\sum_k \gamma_k = 1$; $\sum_l \gamma_{kl} = 0$ for $k = 1,2,3$; and $\sum_k \delta_{ik} = 0$ for $i = 1,2$).[10] From the parameter estimates, ray scale economies can be calculated using the formula:

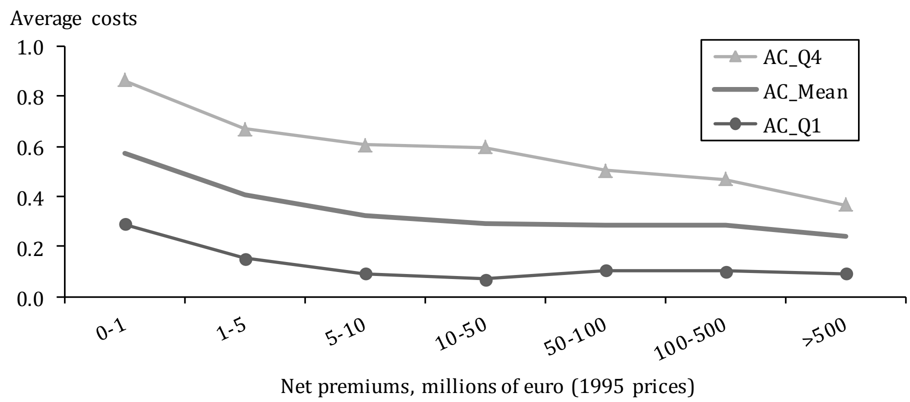
[9] For notational simplicity, time and firm subscripts are dropped.
[10] We have verified that the restriction of linear homogeneity in input prices does not materially affect the results. Unconstrained regression results are available upon request.

$$RSCE = \sum_{i=1}^{2} \frac{\delta \ln \hat{C}(\cdot)}{\delta \ln Y_i}, \tag{4.2}$$

where $\ln\hat{C}(\cdot)$ represents the estimated TCF. A finding of $RSCE < 1$ indicates scale economies, while $RSCE > 1$ indicates diseconomies.

To determine the X-efficiency of Dutch non-life insurers, Berger and Humphrey's (1991) thick frontier approach (TFA) is adopted. We prefer the TFA to other frontier approaches – such as data envelopment analysis (DEA) and the stochastic frontier approach (SFA) – as the degree of cost dispersion in our sample is substantial. Figure 4.2 plots average cost levels (defined as costs over net premiums) for different size classes and cost quartiles (cf. Figure 1 in Berger and Humphrey 1991). The curves AC_Q1, AC_Mean and AC_Q4 represent the mean average cost for the lowest cost quartile, the overall mean average cost and the mean average cost for the highest cost quartile, respectively. Note that the cost quartiles were formed after the size classes were constructed, to make sure that all sizes of insurers are reasonably represented across the cost quartiles. The figure clearly shows that cost dispersion is quite substantial, even within size classes. Firms in the highest cost quartile have average costs levels that are generally more than four times as high as those of firms in the lowest cost quartile. Besides large cost dispersion, Figure 4.2 also shows decreasing average costs with increasing firm size, suggesting economies of scale (*H4.1*).

Figure 4.2. Average costs by size class and cost quartile (1995-2005)



Notes: 'AC_Q1', 'AC_Mean' and 'AC_Q4' are, respectively, the mean average cost of the firms in the lowest cost quartile, the overall mean average cost and the mean average cost of the firms in the highest cost quartile. Average costs are defined as total costs divided by net premiums (output measure).

Under the TFA, Equation (4.1) is estimated separately for firms in the lowest cost quartile (Q1) and firms in the highest cost quartile (Q4). The key assumptions of this approach is that for firms within Q1 and Q4, error term ε reflects only random error, whereas cost differences between firms in Q1 and Q4 are the result of inefficiencies or market factors. Though this assumption is rather ad hoc, it is more intuitive than the assumptions used in many other frontier methods. Of the parametric efficiency studies covered by Berger and Humphrey (1997), more than 25% use the TFA.

The TFA is applied as follows. Let $\ln\hat{C}^{\mathrm{Q}i}(\cdot)$ be the estimated TCF with the parameter estimates of Equation (4.1) obtained when using cost quartile Qi data. As in Figure 4.2, cost quartiles are formed after dividing the sample into seven size classes, $k = 1,2,\ldots,7$. Define the predicted average costs for insurers in cost quartile Qi and size class k, $\widehat{AC}_k^{\mathrm{Q}i}$, as $\widehat{AC}_k^{\mathrm{Q}i} = \hat{C}^{\mathrm{Q}i}(\bar{X}_k^{\mathrm{Q}i})/\overline{NP}_k^{\mathrm{Q}i}$, where $\bar{X}_k^{\mathrm{Q}i}$ is the vector of regressor means (for quartile Qi and size class k) and $\overline{NP}_k^{\mathrm{Q}i}$ represents the mean net premiums (ditto).[11] Predicted costs are divided by net premiums as the latter is our principal proxy for non-life insurance output. Using the predicted average costs for the highest and lowest cost quartile, the proportional cost difference for size class k, $Diff_k$, is given by:

$$Diff_k = \frac{\widehat{AC}_k^{\mathrm{Q4}} - \widehat{AC}_k^{\mathrm{Q1}}}{\widehat{AC}_k^{\mathrm{Q1}}}. \tag{4.3}$$

Equation (4.3) can be decomposed into a part that can be attributed to exogenous market factors, $Market_k$, and a part that cannot, the inefficiency part $Ineff_k$. Say $\widehat{AC}_k^{\mathrm{Q4}^*}$ represents the hypothetical average costs for class k insurers in quartile Q4 if these firms had used the efficient Q1-technology). Then, following Berger and Humphrey (1991), we have

$$Market_k = \frac{\widehat{AC}_k^{\mathrm{Q4}^*} - \widehat{AC}_k^{\mathrm{Q1}}}{\widehat{AC}_k^{\mathrm{Q1}}}, \tag{4.4}$$

and

$$Ineff_k = Diff_k - Market_k = \frac{\widehat{AC}_k^{\mathrm{Q4}} - \widehat{AC}_k^{\mathrm{Q4}^*}}{\widehat{AC}_k^{\mathrm{Q1}}}. \tag{4.5}$$

[11] The predicted costs are computed by taking the means of the regressors, logging the means of the regressors that appear logged in Equation (4.1), computing the cross-products and multiplying by the estimated coefficients. The predicted log costs are transformed back to level costs using the smearing estimator of Duan (1983).

So $Ineff$ captures only the unexplained difference in the estimated cost functions, holding the data constant at Q4. Note that when size class subscript k in Equations (4.3), (4.4) and (4.5) is replaced by time subscript t, the TFA can be used to evaluate X-efficiency developments over time.

4.4 EMPIRICAL RESULTS

We estimate Equation (4.1) under the assumption that all firms, both affiliated and unaffiliated, operate independently from each other. This is fairly in line with reality where subsidiary companies generally operate entirely or highly independently.[12] The full set of parameter estimates is given in Table 4.A.1 in Appendix 4.A.[13] This section presents both conventional estimation results – that are based on the sample of all firms – and thick frontier estimation results – that are based on the Q1 subsample of efficient firms. Berger and Humphrey (1991) argue that scale economies are preferably measured with the thick frontier method and not with the conventional method. They reason that the presence of inefficiencies may bias measures of scale economies obtained via the conventional method. That being so, using the TFA to determine scale economies has its disadvantages as well. Since the TFA estimates are based on only 25% of the observations, 75% of the observations are simply discarded from the analysis. While there may be good reasons to exclude these firms, the reduction in sample size surely reduces the accuracy of the scale economy estimates.

Before going into the results, we note the TCF fits the data quite well. In terms of the adjusted R^2, the variation in total costs is explained better with net premiums as one of the output measures (93%) than with net losses (89%). In an alternative model variant using both proxies of output simultaneously with total investments (not shown here), almost all explanation stems from premiums and total investments, whereas losses play only a very minor role. Apparently, premiums are better output proxies than losses, which we attribute to the latter's stochastic nature. Consequently, in the remainder of this chapter we focus on the premium results.

[12] In alternative estimations (not shown here), we drop this premise of independent operations and, alternatively, assume that cooperation between affiliated firms within groups dominates. The scale economy conclusions were not materially affected.
[13] Since the two logged output variables $\ln Y_1$ and $\ln Y_2$ are highly correlated (both for premiums and losses the pair-wise correlation with total investments exceeds 0.8), we have also estimated Equation (4.1) by using only one output (Y_1). While this improves the accuracy of the estimates, the results are qualitatively the same.

4.4.1 Scale economies and X-efficiency

We start with our hypotheses related to scale economies. Table 4.3 presents estimates of the cost elasticity of output or ray scale economies ($RSCE$) by size class. The first two columns of estimates have been obtained with the conventional method, i.e. OLS estimation of Equation (4.1) using the entire data set. The estimates in the final column have been derived via the thick frontier approach which uses only Q1 data. For comparison reasons, the thick frontier estimates and the conventional estimates are both evaluated at the full sample size class means. When evaluated at the Q1 size class means the thick frontier estimates are quite similar, though.

Table 4.3. Estimated cost elasticities of output by size class (1995-2005)

Net premium size classes (€ millions, 1995 prices)	Cumulative distribution of firms (%)	Conventional method (using full sample)		Thick frontier approach (using Q1 subsample)
		Y_1= Net premiums	Y_1= Net losses	Y_1 = Net premiums
0-1	20	0.80**	0.61**	0.54**
1-5	43	0.83**	0.70**	0.69**
5-10	52	0.87**	0.77**	0.82**
10-50	73	0.91**	0.86**	0.96
50-100	85	0.94**	0.91**	1.08*
100-500	98	0.97	0.99	1.19**
>500	100	1.02	1.07**	1.34**
Overall mean		*0.93***	*0.91***	*1.09*

Notes: 'Y_1= …' indicates that either net premiums or net losses is used as one of the output measures. Cost-elasticity estimates are evaluated at the overall size class means. Confidence bands around the point estimates have been calculated from the estimated variance-covariance matrix. ** and * denote significantly different from one (two-sided) at the 99% and 95% confidence level, respectively.

In agreement with *H4.1*, most Dutch non-life insurance firms seem to operate under increasing returns to scale. Using the conventional method, we find that firms with less than €100 million net premium income (1995 prices) generally face significant unused scale economies. The cumulative size distribution in the second column shows this holds for 85% of the firms in our sample. For firms with an output in the range €100-500 million (1995 prices) constant returns to scale cannot be rejected statistically. According to the thick frontier estimates, scale economies decrease faster with firm size. Consequently, the optimal size– measured roughly at €50 million – is lower when the TFA is used. As can be seen from the last column, the TFA results suggest that for firms with net premium income in the range of €10-50 million (1995 prices) constant returns to scale cannot be rejected. This finding tallies with the average cost curve for Q1 firms in Figure 4.2, which indeed seems to bottom out at the €10-50 million size class. Still, also under the TFA, the majority of firms operate under increasing returns to scale. Finally, both the conventional estimation results (at least when net losses are used as output measure) and the TFA results suggest that the largest non-life insurers face scale diseconomies. These findings are in line with earlier non-life insurance studies that – although using various approaches – generally report economies of scale for small and medium-sized firms and diseconomies of scale for large firms.[14] Bikker and Van Leuvensteijn (2008) observe similar scale economies for the Dutch life insurance industry.

Investigating scale economies over time, Table 4.4 shows estimated cost elasticities by year, evaluated for the average and median firm. The left-hand panel of the table presents the conventional cost function results; the right-hand columns present the thick frontier results. For the average firm the conventional method points to relatively constant unused scale economies of around 7%. Though average firm size increased between 1995 and 2005, the estimated scale efficiency is more or less constant as the estimated coefficients on the quadratic output terms are quite small (see Table 4.A.1 in the appendix). The thick frontier results suggest the average firm continuously operated *above* its optimal scale. For the median firm we find relatively large unused scale economies (14% for both methods) which are increasing as opposed to decreasing. This is caused by a decreasing median firm size. So on the whole our results suggest that scale efficiency in the Dutch non-life insurance industry has *not* improved over our sample period, but rather worsened. This somewhat surprising result contrasts with *H4.2*. While consolidation has led to an increase in average firm size, most firms continue to operate below their optimal scale.

The TFA approach also provides us with X-efficiency estimates of the Dutch non-life insurance industry. Table 4.5 presents the difference in predicted average costs for the highest-cost and lowest-cost quartiles (*Diff*) as well as its decomposition along the lines of Equations (4.4) and (4.5). We observe that for all size classes, market factors only explain a small fraction of the large cost differences (upper panel). Accordingly, the TFA estimates suggest that X-inefficiency is large: the overall inefficiency residual approaches 500%. Note however that the exact magnitude of the inefficiency residual should be treated with due care. The TFA relies on ad-hoc assumptions to disentangle inefficiencies from random cost fluctuations (as do all efficient frontier methods) and cost differences may also reflect unobserved differences in quality, type of products, type of markets, etc. Also our price proxies are far from optimal, as these are not firm-specific. Even so, the measured inefficiency residual is so large – it points to

[14] Suret (1991) reports economies of scale for the Canadian property and casualty insurance industry, whereas Fecher, Perelman, and Pestieau (1991) find modest economies of scale in the French non-life industry of 4%. Hirao and Inoue (2004) examine the Japanese non-life insurance industry and find significant economies of scale for Japanese insurers. Cummins and Weiss (1993) show for the US Property-Liability (P-L) insurance market that small and medium-sized firms are characterized by substantial economies of scale, in the range of 20%, while large firms exhibit mild scale diseconomies of 7%. Their result is broadly replicated by Hanweck and Hogan (1996) who also find that small US P-L firms face economies of scale and large firms experience diseconomies of scale. Cummins and Rubio-Misas (2006) investigate the insurance operations of Spanish firms and similarly conclude that scale economies largely disappear for firms in the largest size quartile. Toivanen (1997), however, reports diseconomies at the firm level for the Finnish non-life insurance industry.

cost X-inefficiencies of 83%, which is also high from an international perspective (see Cummins and Weiss 2000) – that it seems safe to conclude that X-efficiency of Dutch insurance firms is fairly low.

Table 4.4. Estimated cost elasticities by year, evaluated for the average and median firm

Year	Conventional method (using the entire sample)		Thick frontier approach (using the Q1 sub sample)	
	Average firm	Median firm	Average firm	Median firm
1995	0.94 *	0.89 **	1.12*	0.99
1996	0.93 **	0.87 **	1.07**	0.88*
1997	0.94 **	0.87 **	1.11*	0.90*
1998	0.92 **	0.86 **	1.05	0.85**
1999	0.93 **	0.86 **	1.07	0.85**
2000	0.95 *	0.88 **	1.15**	0.92
2001	0.92 **	0.85 **	1.06	0.83**
2002	0.93 **	0.85 **	1.08	0.83**
2003	0.93 **	0.85 **	1.05	0.81**
2004	0.93 **	0.86 **	1.06	0.80**
2005	0.94 *	0.86 **	1.09	0.81**
Overall	*0.93 ***	*0.86 ***	*1.09*	*0.86***

Notes: Net premiums and total investments are used as output measures. ** and * denote significantly different from one (two-sided) at the 99% and 95% confidence level, respectively.

The lower panel of Table 4.5 shows developments in cost differences and the inefficiency residual over time. As expected, and in line with *H4.3,* the estimates suggest that the X-efficiency of Dutch non-life insurers has increased over our sample period, though only moderately so. Inefficiencies have remained substantial however. The inefficiency residual implies cost X-inefficiencies of 78% in 2005, the last year in our sample.

Table 4.5. Decomposition of costs differences by size class and year (in %)

Net premium size classes (€ millions, 1995 prices)	Diff (Eq. 4.3)	Market (Eq. 4.4)	Ineff (Eq. 4.5)	X-inefficiency
0-1	248	19	228	70
1-5	446	36	410	80
5-10	675	35	640	86
10-50	726	45	681	87
50-100	578	54	524	84
100-500	384	35	349	78
>500	250	61	189	65
Overall mean	*530*	*33*	*497*	*83*
Year				
1995	477	45	432	81
2000	393	40	353	78
2005	378	26	352	78

Note: Net premiums and total investments are used as output measures.

4.4.2 Organizational form and focus

Table 4.6 presents parameter estimates that are related to organizational form and focus (again, complete estimation results are in Table 4.A.1 in Appendix 4.A). Only the full-sample estimates are shown. We are more interested in cost differences between stock and mutual insurers in general than in cost differences within the subsample of efficient firms. Indeed, the cost-efficient firms from the Q1 sample are expected to have adequate organizational form and focus.

Table 4.6. Dummy estimates organizational form and focus, conventional method (1995-2005)

	Y_1= Net premiums		Y_1 = Net losses	
Variable	Parameter	s.e.	Parameter	s.e.
STOCK	0.27 **	0.09	0.56 **	0.11
STOCK LOB_{fire}	-0.40 **	0.11	-0.73 **	0.13
STOCK LOB_{health}	0.36 **	0.11	0.14	0.12
STOCK LOB_{motor}	-0.13	0.11	-0.37 **	0.12
STOCK $LOB_{transport}$	-0.59 **	0.20	-0.97 **	0.20
HHI	-0.66 **	0.08	-0.86 **	0.10

Notes: 'Y_1= …' indicates that either net premiums or net losses is used as one of the output measures. *STOCK* is a dummy which takes the value 1 for stock-owned firms and 0 otherwise. *LOB* variables are line-of-business dummies, which take the value 1 if the firm is predominantly active in the respective line of business. *HHI* measures line-of-business specialization. 's.e.' columns give standard errors that are corrected for heteroskedasticity using the sandwich estimator of Hubert (1967) and White (1980). ** and * denote significantly different from zero at the 99% and 95% confidence level, respectively.

The difference in cost levels between stock and mutual insurers is a function of coefficients of the *STOCK* dummy and the four cross terms of *STOCK* and the lines-of-business dummies. Evaluated at the stock sample means of the lines-of-business dummies and using the estimated variance-covariance matrix, we obtain significantly higher average costs for stock insurers. Hence, from these dummy estimates we conclude that stock insurers have, on average, significantly higher cost levels than mutual insurers, which contradicts the expense preference hypothesis (*H4.5*). Similar results have been found by Cummins et al. (2004) and Fecher, Kessler, Perelman, and Pestieau (1993) for the Spanish and French insurance industry, respectively. The estimated coefficients on the interactions between *STOCK* and LOB_q illustrate that cost differences between stock and mutual insurers vary significantly from line to line. The cost advantage of mutual insurers is largest in accident and health insurance, which happens to be their most successful line of business (see Table 4.1). Accordingly, we find additional evidence here consistent with the efficient-structure hypothesis (*H4.4*). Cummins et al. (2004) also find empirical support for the efficient-structure hypothesis. Regarding focus, the estimates in row *HHI* show that more specialized insurers have significantly lower costs, in line with the strategic focus hypothesis (*H4.6*).[15] Cummins and Nini (2002), and Liebenberg and Sommer (2008) also report findings consistent with the strategic-focus hypothesis. Meador, Ryan, and Schellhorn (2000) on the other hand conclude that diversified insurers are more efficient.

4.5 CONCLUSION

The deregulation of Europe's insurance markets has fostered cross-border competition, raising the expectation that competition and, hence, efficiency in the EU Member States has improved over time. Though the new regulatory regime that was adopted across Europe in 1994 was fairly similar to that of the Netherlands, considerable restructuring has occurred since then. Major trends have been consolidation, increased focus and a deteriorating market share of mutual insurers.

This chapter investigates the restructuring of the Dutch non-life insurance industry from a cost-efficiency perspective. As consolidation may have been driven by scale economies, we investigate whether non-life insurance operations are characterized by increasing returns to scale. Using a translog cost model and considering the entire sample of non-life insurance firms, we observe that substantial unused scale economies exist for 85% of the firms at the lower end of the size distribution. For the larger firms we cannot reject constant returns to scale. When we measure scale efficiency for the subset of the 25% most efficient firms only, we find a steeper decrease of scale economies with firm size: the majority of firms, the smaller ones, operate under economies of scale but the largest firms face diseconomies of scale. Remarkably, our results suggest that scale efficiency has *not* improved over our sample period, but rather worsened.

Besides scale economies, X-efficiency differences between insurers also provide an incentive for M&A. When relatively efficient firms take over less efficient ones, consolidation is hypothesized to increase the X-efficiency in the industry. Using the thick frontier approach, we observe a substantial degree of X-inefficiency in the Dutch non-life insurance industry, though this may partly reflect unobserved differences in quality, products and markets. Consistent with expectations, X-efficiency has moderately improved over time.

The cost implications of organizational form and focus are also examined. Two organizational form hypotheses are tested: the expense-preference hypothesis and the efficient-structure hypothesis. The former hypothesis predicts that mutual insurers are generally less cost-efficient than stock insurers, as the mutual ownership form provides weaker mechanisms for controlling owner-manager agency conflicts. The efficient-structure hypothesis states that

[15] Note that the TFA yields similar signs on most dummies, yet the estimated standard errors are much larger. This is the result of lower variation in the dummy variables in the Q1 sample set.

stock and mutual insurers are relatively successful in lines of business where they have comparative advantages. In agreement with most of the respective literature, our results contradict the expense-preference hypothesis. Further, cost differences between stock and mutual insurers are found to vary significantly across the different lines of business, whereas the cost advantage of mutual insurers is largest in accident and health insurance. This happens to be the most successful line of business for the mutual ownership form − in terms of market share − confirming the efficient-structure hypothesis. Finally, regarding focus, our estimates show that more specialized insurers have significantly lower costs, in line with the strategic-focus hypothesis.

APPENDIX 4.A PARAMETER ESTIMATES

Table 4.A.1. Parameter estimates of translog cost models (1995-2005)

	Conventional method (using the full sample)				Thick frontier approach (using Q1 subsample)	
	Y_1 = Net premiums		Y_1 = Net losses		Y_1 = Net premiums	
Variable	parameter	s.e.	Parameter	s.e.	parameter	s.e.
$\ln Y_1$	0.79**	0.03	0.64**	0.03	1.02**	0.05
$\ln Y_2$	0.14**	0.03	0.27**	0.03	0.06	0.05
$\ln P_1$	0.83**	0.08	0.89**	0.09	0.44**	0.16
$\ln P_2$	0.07	0.06	0.01	0.07	0.01	0.14
$\ln P_3$	0.11	0.10	0.10	0.13	0.55	0.22
$\ln^2 Y_1$ [a]	0.00	0.01	0.02	0.01	0.05**	0.01
$\ln^2 Y_2$ [a]	0.00	0.00	0.01	0.01	0.00	0.01
$\ln^2 P_1$ [a]	-0.37	0.42	-0.39	0.52	0.72	1.06
$\ln^2 P_2$ [a]	0.07	0.23	0.24	0.28	-0.09	0.58
$\ln^2 P_3$ [a]	0.33	0.43	0.52	0.53	-0.30	1.00
$\ln Y_1 \ln Y_2$ [a]	0.01	0.01	0.00	0.01	0.01	0.01
$\ln Y_1 \ln P_1$ [a]	0.03	0.05	0.01	0.06	-0.01	0.07
$\ln Y_1 \ln P_2$ [a]	0.05	0.05	0.05	0.05	0.03	0.08
$\ln Y_1 \ln P_3$ [a]	-0.08	0.07	-0.07	0.08	-0.02	0.11
$\ln Y_2 \ln P_1$ [a]	-0.06	0.06	-0.04	0.06	-0.11	0.08
$\ln Y_2 \ln P_2$ [a]	-0.07	0.05	-0.10	0.06	-0.08	0.09
$\ln Y_2 \ln P_3$ [a]	0.13	0.07	0.14	0.09	0.20	0.11
$\ln P_1 \ln P_2$ [a]	0.31	0.27	0.34	0.33	-0.47	0.64
$\ln P_1 \ln P_3$ [a]	0.06	0.26	0.06	0.32	-0.26	0.62
$\ln P_2 \ln P_3$ [a]	-0.38	0.44	-0.58	0.54	0.55	1.08
HHI	-0.66**	0.08	-0.86**	0.10	-0.19	0.22
STOCK	0.27**	0.09	0.56**	0.11	-0.08	0.15
LOB_{fire}	0.37**	0.08	0.70**	0.09	0.12	0.14
LOB_{health}	-0.76**	0.09	-0.51**	0.10	-0.51**	0.14
LOB_{motor}	0.05	0.10	0.25*	0.11	0.17	0.15
$LOB_{transport}$	0.06	0.09	0.19	0.11	0.26	0.20
$STOCK\ LOB_{fire}$	-0.40**	0.11	-0.73**	0.13	-0.48	0.27
$STOCK\ LOB_{health}$	0.36**	0.11	0.14	0.12	0.41*	0.16
$STOCK\ LOB_{motor}$	-0.13	0.11	-0.37**	0.12	0.06	0.19
$STOCK\ LOB_{transport}$	-0.59**	0.20	-0.97**	0.20	-0.52	0.30
GROUP	0.02	0.03	0.04	0.04	0.02	0.08
T	0.01	0.01	0.01	0.01	0.01	0.02
Constant	-3.93**	0.30	-3.76**	0.36	-5.46**	0.62
# observations	2,148		2,148		535	
R^2, adjusted	0.93		0.89		0.90	

[a] Output and price variables are in deviation from their full sample means. Consequently, estimated average scale economies can be easily conferred from the estimates of the linear output terms' coefficients. *Notes*: 'Y_1 = …' indicates that either net premiums or net losses is used as one of the output measures. Variables Y_2, P_1, P_2 and P_3 represent total investments, price of labor, price of equity capital and price of debt capital, respectively. *HHI* measures line-of-business specialization. *STOCK* is a dummy which takes the value 1 for stock firms and 0 otherwise. *LOB* variables are line-of-business dummies that take the value 1 if the firm is predominantly active in the respective line of business. Dummy *GROUP* (1 for firms affiliated to a group) is included to control for group affiliation. *T* is a linear time trend. 's.e.' columns give standard errors that are corrected for heteroskedasticity using the sandwich estimator of Hubert (1967) and White (1980). ** and * denote significantly different from zero at the 99% and 95% confidence level, respectively. All monetary variables have been deflated to 1995 prices.

Chapter 5

INVESTMENT RISK-TAKING INCENTIVES[*]

[*] This chapter is based on Gorter and Bikker (2013).

5.1 INTRODUCTION

Investment behavior of pension funds and insurance firms, particularly their risk-return preferences, is of great importance. While more investment risk typically results in higher expected returns, it also tends to increase their asset-liability mismatch, thereby endangering future pension and insurance benefits. Recent crises in the financial markets have made the potential adverse consequences of institutional investment risk taking crystal clear. Indeed, the financial positions of defined benefit (DB) pension funds all over the world have been eroded. Buffers of insurance firms have also been affected, yet materially less so. While it is widely known that in practice DB pension funds tend to run a larger mismatch risk than insurance firms (see, e.g., Broeders, Chen, and Koos 2011), empirical research into the root cause of this stylized fact is, to the best of our knowledge, yet absent. This chapter aims to fill part of this void in the literature by comparing investment risk taking of DB pension funds, life insurers and non-life insurers in the Netherlands.

Theory offers two main hypotheses on investors' risk appetite. According to the risk-management hypothesis, financially constrained investors have an incentive to invest prudently because of bankruptcy costs (Smith and Stulz 1985) and the potential inability to accept profitable future investment projects (Mayers and Smith 1987). Sommer (1996) and Cummins and Danzon (1997) present empirical evidence that insurance firms face financial distress costs which limits their risk taking. Specifically these authors show that insurance is priced as risky debt, and that insurance prices are inversely related to insurer default probabilities. Hence, taking more investment risk comes at a cost of lower insurance policy profit margins. Additional risk management incentives are provided for by insurance regulation. Indeed, when capital falls below the regulatory minimum, the prudential supervisor, which in the Netherlands is DNB, assumes control over the respective insurer. As a result policyholders are protected against extensive losses and owners and management have an incentive to stay away from the regulatory minimum.

Risk-management incentives are also expected to be relevant for pension funds. Occupational pension funds are principally funded by employers that are likely to have risk-management incentives. In fact, Rauh (2009) concludes for US DB pension plans that risk-management incentives dominate investment behavior. That being so, in comparison to insurers, we expect the risk-management incentives of pension funds to be more subdued, as pension funds do not face financial distress costs. Pension funds are trusts, and when assets fall below liabilities, a fund does not go bankrupt, employees are not laid off and non-marketable assets are not lost.

The other leading hypothesis about investor risk taking is the risk-shifting hypothesis, which states that investors have an incentive to invest in risky securities, especially when in financial distress (Jensen and Meckling 1976). The incentive for risk shifting originates from an asymmetry of pay-offs: stakeholders in institutional investors that benefit from the upside of more investment risk and face limited downside may have an incentive to increase portfolio risk. Because of limited liability, stock owners have such asymmetric payoffs. Accordingly, the ownership-structure hypothesis predicts that stock-owned insurers have stronger risk taking incentives than their mutual peers (Lee, Mayers, and Smith 1997). The underlying logic is that stock owners are able to increase the value of their claims by increasing portfolio risk while mutual owners typically are not, as mutual owners are by definition also policyholders. Consequently, in case of mutual insurers, the benefits from risk shifting are low and more diffuse (Esty 1997).

As the organizational form of pension funds appears more comparable to that of mutual insurers than to that of stock insurers, one may expect that risk-shifting incentives are relatively less important in the pension domain. However, occupational pension funds have company sponsors that could engage in risk-shifting behavior vis-à-vis their pension plan participants. Risk-shifting incentives may be especially relevant when sponsors are not fully liable for shortfalls in the pension fund, which happens to be the case in the Netherlands. But even when a sponsoring company is fully liable, limited liability restricts the downside risk for its owners. This then creates an incentive for the sponsor to favor a risky investment strategy when financial conditions deteriorate (Sharpe 1976, Treynor 1977).

This chapter assesses the relevance of risk-management and risk-shifting incentives for Dutch institutional investors over the fifteen year period 1995-2009. Our data set covers DB pension funds, life insurers and non-life insurers that collectively manage over €1 trillion in assets (2009 figure), which is almost twice Dutch GDP. The chapter is related to the empirical literature on the investment policies of pension funds (e.g. Lucas and Zeldes 2009, Rauh 2009, Bikker, Broeders, and De Dreu 2010, Bikker, Broeders, Hollanders, and Ponds 2012) and insurance firms (e.g. Cummins and Sommer 1996, Baranoff and Sager 2002, 2003). Note that the Dutch setting is particularly suitable to examine investment behavior, since institutional investors in the Netherlands are, in principle, free to choose their desired risk return trade-off, and allocate their assets accordingly. In many other countries, in contrast, investors face quantitative restrictions to limit risk taking.[1]

A natural and often-used measure for investment risk taking is the equity allocation, i.e. the percentage of equities in the investment portfolio (see, for instance, Lee et al. 1997, Rauh 2009).

[1] In a study of the regulation of institutional investors in the major OECD countries, Davis (2002) reports that only the UK and the Netherlands do not impose quantitative restrictions on equity holdings for life insurers.

Indeed, equity investments expose investors to considerable income and balance sheet volatility. The main drawback of the equity allocation as a risk proxy is, however, that it ignores interest rate risk. While this drawback would bias the analysis when there is a negative correlation between equity and interest rate risk taking, the opposite appears to be the case. Dutch insurers and pension funds with a large interest rate risk exposure also tend to have a large equity risk exposure, and vice versa. Accordingly, the equity allocation appears to be a suitable proxy for investment risk taking. We examine equity allocations both across investors and over time. In the cross-sectional analysis we focus on differences in the level of risk taking across institutional investors. In the time-series analysis, we examine how trading behavior responds to past returns, that is, we examine feedback trading. While there is a sizeable literature on feedback trading, the microprudential risk implications of such behavior have typically been passed over. It can easily be shown, however, that if equity prices follow a random walk, buying equities when equities are down (i.e. negative feedback trading, or rebalancing) is risky. Under the same definition, positive feedback trading can also be risky, that is, in upward markets. Hence, feedback trading provides interesting insights into investors' risk taking over time.

The chapter is organized as follows. The next section presents further background to the risk-taking behavior of institutional investors and introduces testable hypotheses. Section 5.3 explains the used methodologies, describes the data set and provides summary statistics and information about developments over time. Section 5.4 shows empirical results, both for the comparison of risk taking across types of institutional investors and for the comparison of investment risk taking over time. The last section concludes.

5.2 HYPOTHESES

5.2.1 Risk-bearing capacity

As mentioned in the introduction, theory provides two competing hypotheses about the relationship between risk-bearing capacity and risk taking: the risk-management hypothesis and the risk-shifting hypothesis. Given that Dutch insurers generally hold capital levels well in excess of the regulatory requirements (De Haan and Kakes 2010), and given Rauh's (2009) finding that risk-management incentives dominate risk-shifting incentives for US DB pension plans, we expect, on average, a positive relationship between risk-bearing capacity and investment risk taking.

H5.1: Investors with more risk-bearing capacity take more investment risk, and vice versa.

We use several empirical measures of investor's risk-bearing capacity. The first is the capital ratio, which provides a direct insight into risk-bearing capacity.[2] With more capital, an institution can shoulder riskier asset portfolios and/or provide safer returns to its stakeholders (Gatzert and Schmeiser 2008). Along the lines of *H5.1*, we expect institutional investors with more capital to take more investment risk, and vice versa.

H5.1a: Institutional investors with more capital take more investment risk, and vice versa.

Another risk factor is size. Larger firms generally have more diversification benefits, both on the asset- and the liability-side of their balance sheets. As diversification reduces an investor's overall risk profile, larger investor are expected to have more risk-bearing capacity. Accordingly, we expect a positive relationship between investment risk taking and size. Larger firms also benefit from scale economies, in being able to set up a more sophisticated risk-management organization. Paradoxically, more intensive risk management often leads to more risk taking, as it allows financial institutions to measure their risks more accurately and deploy their scarce capital in the supposedly most efficient way. Large firms may also suffer from overconfidence when they put too much trust in (self-developed) theories and models. Hence, we expect larger firms to take more investment risk, though it is not entirely clear whether this is due to greater risk-bearing capacity or due to less risk aversion.

H5.1b: Larger institutional investors take more investment risk, and vice versa.

The insurance industry consists of various lines of business with diverging risk profiles. We anticipate that in more volatile lines of business, insurers are less eager to take investment risk. An interesting finding from the Solvency II[3] calibration study QIS5 is that underwriting risk is typically larger for non-life than for life insurers (EIOPA 2011). Consequently, life insurers can allocate more capital to market risk. In that context, it is important to differentiate between traditional life insurance, where the insurer bears the investment risk, and unit-linked life insurance, where the investment risk is primarily borne by the policyholders. In view of that,

---

[2] In this chapter, the capital ratio is an important proxy of risk-bearing capacity. The capital ratio is defined as net asset value (i.e. assets minus liabilities) to total assets. While common in banking, the capital ratio is not the typical solvency indicator in the insurance and pensions industry. In the insurance industry, solvency conditions are usually presented in terms of the solvency ratio. The solvency ratio equals the actual solvency margin divided by the required solvency margin. In the context of pension funds, solvency conditions are typically presented in terms of the funding ratio, which is the ratio of total assets to total liabilities. For ease of comparison, however, we use capital ratios for both insurers and pension funds.

[3] Solvency II is the envisaged new European supervisory framework for insurance firms that will introduce, among other things, risk-based solvency requirements.

Dutch insurance regulation puts a lower capital charge on unit-linked technical provisions (De Haan and Kakes 2010).

As explained in Chapter 4, Dutch non-life insurance consists of five lines of business: accident and health; motor; marine, transport and aviation; fire and other property risk; and miscellaneous insurance. Miscellaneous insurance includes a wide variety of insurance products such as travel insurance, legal expenses insurance and credit insurance. Some of these lines are more volatile than others. An often-used measure to capture non-life underwriting risk is the standard deviation of the loss ratio (e.g., Meyers 1989, Lamm-Tenant and Starks 1993, Guo and Winter 1997). The loss ratio is the ratio of losses incurred to premiums earned and is a measure for underwriting profitability. Table 5.1 shows percentiles of the distribution of this risk measure for Dutch non-life insurers, both on the line of business and the firm level (based on our data set, described in subsection Subsection 'Data' below). The figures show that motor insurance tends to be a relatively stable line of business in the non-life insurance industry, whilst the category marine, transport and aviation insurance, by contrast, is more than twice as volatile as the reference category motor insurance. In the analyses, we account for differences between the lines of business by including line-of-business dummies (motor insurance acts as the reference group).

Table 5.1. Standard deviation of loss ratio (1995-2009)

	Number of firms	Percentiles		
		p_{25}	p_{50}	p_{75}
Line-of-business level				
Accident and health	107	0.07	0.12	0.21
Motor	62	0.05	0.08	0.16
Marine, transport and aviation	54	0.13	0.18	0.49
Fire and other property risk	126	0.09	0.15	0.26
Miscellaneous insurance	89	0.06	0.11	0.19
Firm level	*199*	*0.07*	*0.12*	*0.19*

Notes: Loss ratio is the ratio of losses incurred to premiums earned. Standard deviations have been calculated for non-life insurers with eight or more consecutive years of data. Figures shown are after selection. See Subsection 'Data' for the selection criteria.

For insurers, reinsurance and group affiliation are also potential determinants of investment behavior. Insurers that cede more business to reinsurers have less underwriting risk and can thus allocate more capital to investment risk. If group control is imperfect, i.e. there is considerable independence on the part of affiliated firms, consistent with Chapter 4, then affiliated firms have an incentive to take more risk.

H5.1c: *Insurers with more underwriting risk take less investment risk, and vice versa.*

For pension funds, we hypothesize that the share of active participants positively influences the degree of risk taking. There are three main reasons to expect such a relationship. First, Sundaresan and Zapatero (1997) and Lucas and Zeldes (2009) argue that equity investing (our measure of risk taking) may hedge against increases in pension benefits. Pension benefits of active participants are determined by real wage developments, which are positively correlated with equity returns. Second, the share of active participants largely determines the effectiveness of raising contributions to stave off underfunding of the pension fund. Third, active participants can accommodate investment losses by working more and/or longer (Bikker et al. 2012).

H5.1d: *Pension funds with more active participants take more investment risk, and vice versa.*

Another pension-fund-specific variable we consider is total pension wealth per plan participant. At the end of the day, pension funds invest for private persons that tend to invest more in equity the larger their savings are (Cohn, Lewellen, Lease, and Schlarbaum 1975).

H5.1e: *Pension funds with wealthier plan participants take more investment risk, and vice versa.*

5.2.2 Pension funds versus insurance firms

Pension funds are trusts that do not go bankrupt and do not face financial distress costs. When assets fall below liabilities, plan participants cannot walk away as they are obliged to take part in the pertaining pension scheme. By contrast, insurers are likely to lose policyholders when solvency capital runs low. In fact, empirical evidence shows that insurers do indeed face financial distress costs. Sommer (1996) and Cummins and Danzon (1997) show insurance prices are inversely related to insurer default probabilities. Consequently, we expect pension funds to take more investment risk than insurance firms.

Another, related reason why pension funds are expected to take more investment risk concerns solvency regulation. When an insurer’s capital ratio falls below the regulatory minimum, DNB – the prudential supervisor in the Netherlands – assumes control over the insurer. Regulation is different for pension funds. When a pension fund falls below the minimal capital ratio of roughly 5%, it is normally given three years to recover. This may be done through (a combination of) reducing or eliminating indexation, increasing contributions, receiving a subordinate loan or by renegotiating the unconditional rights between the employer and trade unions. In exceptional circumstances, the government can allow a longer recovery period. This happened in 2009, in fact, when the government extended the recovery period to five years. Consequently, we expect pension funds to take more investment risk than insurance firms.

H5.2: *Pension funds take more investment risk than insurers.*

H5.3: *The relationship between capital and investment risk taking is more pronounced for insurers than for pension funds.*

We anticipate *H5.2* and *H5.3* to hold, despite the relative risk insensitivity of the current insurance solvency requirements. Though underwriting risk is to some extent reflected in the solvency requirements, investment risk is not (De Haan and Kakes 2010).[4] By contrast, pension funds regulation in the Netherlands has been risk-based since 2007, when the Financial Assessment Framework (in Dutch: Financieel Toetsingskader, FTK) became effective (Bikker and Vlaar 2007). According to this risk-based regulatory framework, pension funds are required to hold a higher risk margin for risky investments such as equities. In principle available capital should be sufficient to maintain a less than 2.5% probability of insolvency over a one-year horizon. For an average pension fund, this implies a target funding ratio of around 130 percent. Since pension funds do not have to de-risk when buffers run low, we expect them to be less risk sensitive in their investment behavior than insurers. Note that the FTK did not only make supervision more risk-based, it also changed the way pension liabilities are calculated. Until end-2006, pension liabilities were calculated on the basis of a fixed actuarial interest rate of at most 4% (a figure that had not changed since 1969), which was considered a conservative estimate of the expected long-run real return on the asset portfolio. With the introduction of the FTK, the fixed actuarial interest rate to calculate liabilities was replaced by the (nominal) market rate. This was an important change that had an immediate impact on capital ratios, or – in the pension domain – funding ratios.

5.2.3 Stock versus mutual insurers

The stock and mutual organizational form are present in insurance markets around the world (Swiss Re 1999). In Chapter 4 we have investigated the cost-efficiency implications of organizational form. In this chapter we focus on the risk-implications.

Lamm-Tenant and Starks (1993) investigate risk taking of stock and mutual insurers in the US, yet look at underwriting rather than investment risk. Lamm-Tenant and Starks find that stock insurers bear significantly more underwriting risk than mutual insurers. Lee et al. (1997) examine investment risk taking across stock and mutual insurers. They propose the ownership-

[4] This will change when the European Union will implement Solvency II, the envisaged risk-based regulatory approach for European insurance companies.

structure hypothesis, which predicts that stock-owned firms have stronger risk-taking incentives than mutual insurers. Limited liability enables stock owners to increase the value of their claims by increasing portfolio risk. For mutual owners, risk-shifting incentives are low or non-existent (Esty 1997). Mutual ownership claims are held by policyholders and are principally inalienable. Stock insurers' ownership claims, on the other hand, are held by specialized residual claimants whose claims are alienable. This alienability facilitates stock owners to offer management pay-for-performance compensation packages, including equity holdings and equity options, therewith encouraging them as well to take asset risk (Mayers and Smith 1992).[5]

Lee et al. (1997) study changes in U.S. insurers' asset portfolios after guaranty fund enactment. They find that only stock insurers shifted to riskier assets following fund enactment, providing support for the ownership-structure hypothesis. Since Lee et al. (1997) examine asset risk taking across stock and mutual insurers following an exogenous event, they naturally control for differences in underwriting risk between both types of insurers. Controlling for underwriting risk is important, as stock insurers are both predicted and found to be associated with riskier insurance business (e.g. Fama and Jensen 1983, Smith and Stutzer 1990, Mayers and Smith 1992, 1994, Lamm-Tenant and Starks 1993). Hence, we expect that, corrected for the relevant risk factors, summed up in Section 5.2.1, stock insurers take more investment risk than mutual insurers.

H5.4: *Stock insurers generally take more investment risk than mutual insurers.*

5.2.4 Types of pension funds

In the Netherlands, there are three types of pension funds: company pension funds, industry-wide pension funds and professional group pension funds. While company pension funds are most common, industry-wide pension funds govern more assets. Professional group pension funds manage the pension assets of professionals such as doctors and notaries. A key difference between these pension fund types is that in professional group pension funds, the participants are typically self-employed, while in the other funds the participants are employees of the employer. Hence, while there are potentially risk-shifting incentives in company and industry-wide pension funds, such incentives are absent in professional group pension funds. On this basis, we expect less risk taking by professional group pension funds compared to other types of pension funds.

[5] Though in principle mutual insurers could mimic such compensation packages, to the best of our knowledge this does not happen in practice.

H5.5: Professional group pension funds take less investment risk than company and industry-wide pension funds.

5.3 METHODOLOGY AND DATA

5.3.1 Methodology

We investigate investment risk taking from two angles. First, the cross-sectional variation in equity allocations is analyzed. Subsequently, we turn our focus to market developments and analyze how the various investor types react to equity price changes over time. Since prudent investors are more likely to survive than less-prudent investors, especially during our volatile sample period, one could argue that survivorship bias is likely to affect our results. To gauge potential survivorship bias, we do both balanced and unbalanced estimations. The unbalanced data set includes institutional investors that have ceased operations.

5.3.1.1 Equity allocations across investors

Equity investments typically expose investors to considerable income and balance sheet volatility and have therefore been frequently used to measure risk taking. Lee et al. (1997), Cummins and Nini (2002) and De Haan and Kakes (2010) use this measure for insurers, while Lucas and Zeldes (2009) and Bikker et al. (2012) use this measure in a study of, respectively, US and Dutch pension funds. Unfortunately, the equity allocation does not take interest rate risk into account, which is an important risk for (life) insurers and pension funds. Ideally we would therefore also use an interest rate risk proxy, such as the duration gap or an interest rate risk regulatory capital requirement. Unfortunately such proxies are unavailable over the studied time frame, so we leave a thorough study of interest rate risk taking for future research. We expect our results are robust, though, as data over 2009 show a positive correlation between interest rate risk and equity risk, both for insurers and pension funds.[6]

[6] For life insurers, we have used data from the Solvency II QIS5 study to establish a sample correlation of 0.5 between interest rate risk and equity risk, where both risks are approximated by the respective solvency capital requirements scaled by total investments. For pension funds, we have used comparable data, but then from the Dutch financial assessment framework data, to establish a sample correlation of 0.4 between interest rate risk and equity risk.

Our empirical specification reads as follows:

$$W_{i,t} = \alpha_0 + \alpha_1 CR_{i,t-1} + \alpha_2 \ln SIZE_{i,t-1}$$
$$+D_i^I[\alpha_3 + \alpha_4 CR_{i,t-1} + \alpha_5 \ln SIZE_{i,t-1} + \alpha_6 STOCK_i + \sum \alpha_{7+q} LOB_{i,q,t} + \alpha_{13} G_{i,t} + \alpha_{14} R_{i,t-1}]$$
$$+D_i^{PF}[\alpha_{15} SA_{i,t} + \alpha_{16} \ln WP_{i,t-1} + \alpha_{17} IF_i + \alpha_{18}] + \sum \alpha_{19+t} YD_t + \varepsilon_{i,t}, \quad (5.1)$$

where $W_{i,t}$ is the equity allocation of investor i in year t, written as a linear function f of coefficient vector α and a set of explanatory variables. Dummy variables D_i^I (1 for insurers) and D_i^{PF} (1 for pension funds) allow for differences in the coefficients between insurers and pension funds. $CR_{i,t-1}$ and $\ln SIZE_{i,t-1}$ stand for, respectively, lagged capital ratio and lagged investor size (measured by total investment portfolio size) in logarithms, and are available for both insurance firms and pension funds. We take lags to avoid simultaneity bias. Nonetheless, even the lagged capital ratio is potentially endogenous. All else equal, more equity in the investment portfolio requires an investor to hold more capital to attain the same level of risk, where risk is the probability of insolvency for insurers and the probability of underfunding for pension funds. We address this potential endogeneity problem in our estimation procedure.

Dummy variable $STOCK_i$ takes the value 1 for stock insurers and 0 otherwise, and captures equity-allocation differences between stock and mutual insurers. $LOB_{i,q,t}$ for $q = 0$ to $q = 5$, are insurance line-of-business dummies for six of the seven lines,[7] which take the value 1 if a firm is predominantly active in the respective line.[8] The relatively stable non-life insurance line of business motor insurance acts as the reference group. Dummy variable $G_{i,t}$ (1 for firms affiliated to a group) is included to control for group affiliation. $R_{i,t-1}$ is the reinsurance ratio lagged, i.e. last year's premiums ceded as a percentage of last year's total premiums. This reinsurance proxy is used by, among others, Cummins and Nini (2002) and De Haan and Kakes (2010). $SA_{i,t}$ and $\ln WP_{i,t-1}$ are pension-fund-specific variables, which measure, respectively, the share of active participants and lagged total assets per participant (in logarithms). IF_i and PG_i are dummy

[7] The seven lines-of-business dummies consist of five non-life insurance categories (accident and health, motor, marine, transport and aviation, fire and other property risk, and miscellaneous insurance) and two life insurance lines of business, namely unit-linked life and traditional life.

[8] Note that non-life insurers are not permitted to write life insurance policies, and, vice versa, life insurers are not permitted to write non-life insurance policies. That being so, insurance holding companies are allowed to have both life and non-life subsidiaries. A non-life insurer is predominantly active in a specific non-life insurance line of business when more than 50 percent of its premium income comes from this line. When a specific non-life insurer has multiple lines of business and is not predominantly active in any line of business, the line-of-business dummies are all 0 for this insurer. For life insurers, our data set discerns two lines of business: traditional life insurance and unit-linked life insurance. Given the long-term nature of life insurance, we distinguish between both types of life insurers by reserves than by premium income. A life insurer is predominantly active in traditional life when more than 50 percent of its reserves are traditional life reserves.

variables that are 1 for industry-wide and professional group pension funds, respectively, and 0 otherwise. The reference group are the omitted company pension funds. Year dummies YD_t are included to account for equity-market developments over the sample period. Finally, the error term is represented by $\varepsilon_{i,t}$.

5.3.1.2 Equity allocations over time

The second part of our analysis, on equity allocations over time, touches upon the issue of feedback trading. There is a sizeable empirical literature on feedback trading by institutional investors. This literature distinguishes between positive and negative feedback trading. Positive-feedback traders buy equities when their prices rise, and sell equities when their prices fall. Negative-feedback traders do exactly the opposite: they sell equities when equities are up, and buy equities when they are down. While available studies find that past equity returns significantly influence today's trading behavior, evidence on the sign of such feedback trading is mixed. Grinblatt, Titman, and Wermers (1995) find evidence of positive feedback trading for US institutional investors, yet Lakonishok, Shleifer, and Vishny (1992) and Gompers and Metrick (2001) do not observe significant positive feedback trading. The question then rises, why some institutional investors engage in positive feedback trading and others do not. We argue that risk-bearing capacity probably plays an important role.

Existing studies on Dutch pension funds and insurance firms report negative feedback trading (Kakes 2008, Bikker et al. 2010, De Haan and Kakes 2011). Negative feedback trading, or rebalancing, is often motivated by mean reversion in equity prices. Under mean reversion, an institution can achieve higher returns by buying equities when markets are down and selling equities when markets are up. However, since the equity cycle can be long and volatile (Balvers, Wu, and Gilliland 2000, Spierdijk, Bikker, and Van der Hoek 2012) and the degree of mean reversion is likely to be small and uncertain (Pástor and Stambaugh 2012), rebalancing can be a particularly risky trading strategy. When equity prices fall, risk-bearing capacity of equity investors also falls. Buying equities when risk-bearing capacity is down is risky business and only possible for unconstrained or loosely constrained investors. Constrained investors, on the other hand, can be forced to liquidate their equity holdings just when the majority of investors are trying to shed risk, a phenomenon known as fire sales (Shleifer and Vishny 1992). The reverse happens in upturns. Rising equity markets increase investors' willingness and ability to take risk (Black 1988).

In a related study on Dutch pension funds and insurance firms over the period 1999-2005, De Haan and Kakes (2011) report evidence of negative feedback trading. Other than De Haan and Kakes (2011), who use quarterly survey data that are collected for statistical purposes (balance-of-payments statistics), this study uses regulatory data with an annual frequency. Main advantages are better coverage (all regulated investors are included), a longer time-span (1995-2009), including the recent financial crisis, and a breakdown of traditional life and unit-linked life investments. The latter is important, as the market risk in unit-linked investment portfolios is primarily borne by policyholders and not by the respective life insurers (see Section 5.2.1).

To investigate the relationship between equity trading and equity price changes, we decompose the equity allocation of investor *i* in year *t*, $W_{i,t}$, into three factors:

$$W_{i,t} = \frac{E_{i,t}}{TI_{i,t}} = \frac{E_{i,t}^{REV}}{TI_{i,t}} + \frac{E_{i,t}^{NCF}}{TI_{i,t}} + W_{i,t-1}\frac{TI_{i,t-1}}{TI_{i,t}}, \tag{5.2}$$

where $E_{i,t}^{REV}$ is the euro amount of equity gains or losses. $E_{i,t}^{NCF}$ is the euro amount of net purchases or sales and ratio $TI_{i,t-1}/TI_{i,t}$ is the inverse of portfolio growth. When non-equity assets grow and the value of equity holdings remains constant, the equity allocation drops as a result of positive portfolio growth.

Equation (5.2) is an identity where the factors on the right-hand side of the equation all have a one-to-one effect on the equity allocation. What we are interested in, however, is whether price changes and equity trading are generally reinforcing (evidence of positive-feedback trading) or counterbalancing (evidence of negative-feedback trading). To investigate this, we introduce the following basic model:

$$W_{i,t} - W_{i,t-1}\frac{TI_{i,t-1}}{TI_{i,t}} = \frac{\Delta E_{i,t}}{TI_{i,t}} = \beta_0 + \beta_1\frac{E_{i,t}^{REV}}{TI_{i,t}} + \upsilon_{i,t}, \tag{5.3}$$

with $\Delta E_{i,t}$ representing equity portfolio growth (in euro's) of investor *i* in year *t*. Comparing Equations (5.2) and (5.3), notice we have dropped trading factor $E_{i,t}^{NCF}/TI_{i,t}$, yet added parameters β_0 and β_1, and disturbance term $\upsilon_{i,t}$. Estimating Equation (5.3) by OLS gives for β_1 an estimate

$$b_1 = \frac{\widehat{\text{Cov}}\left(E_{i,t}^{REV}/TI_{i,t},\ \Delta E_{i,t}/TI_{i,t}\right)}{\widehat{\text{Var}}\left(E_{i,t}^{REV}/TI_{i,t}\right)} = 1 + \frac{\widehat{\text{Cov}}\left(E_{i,t}^{REV}/TI_{i,t},\ E_{i,t}^{NCF}/TI_{i,t}\right)}{\widehat{\text{Var}}\left(E_{i,t}^{REV}/TI_{i,t}\right)},$$

where we have used the identity $\Delta E_{i,t} = E_{i,t}^{REV} + E_{i,t}^{NCF}$. Hence, when $b_1 > 1$ ($b_1 < 1$), the sample correlation between market returns and net equity purchases is positive (negative), suggesting positive (negative) feedback trading. Note that feedback trading implies exogenous equity returns, which seems likely as equity returns are mainly determined by the market rather than by individual investors.

In order to allow for asymmetric feedback trading, which is different trading behavior in bear and bull markets, we generalize Equation (5.3) to

$$\frac{\Delta E_{i,t}}{TI_{i,t}} = \beta_0 + D_{i,t}^{POS}\beta_1^{POS}\frac{E_{i,t}^{REV}}{TI_{i,t}} + D_{i,t}^{NEG}\beta_1^{NEG}\frac{E_{i,t}^{REV}}{TI_{i,t}} + \upsilon_{i,t}, \tag{5.4}$$

where $D_{i,t}^{POS}$ and $D_{i,t}^{NEG}$ are investor-specific dummy variables that take the value 1 if equity returns in year t are positive and negative, respectively, and 0 otherwise. Further, we allow for different behavior for the types of institutional investors we consider, by multiplying the β coefficients in Equation (5.4) with sector dummies. This approach is equivalent to estimating Equation (5.4) separately for different investor types.

5.3.2 Data

This analysis draws on a comprehensive regulatory data set from DNB.[9] It covers all regulated pension funds, life insurers and non-life insurers in the Netherlands over the 1995 to 2009 period.[10] Note that for life insurers, we only include investments for their own account, so that unit-linked investments are excluded from the equity allocations used in estimations of Equation (5.1).[11] After sanitization, the data set contains 12,799 institution-year observations.[12] Note that for pension funds, transaction and revaluation data are only available from 2000 and onwards. This holds also for strategic equity allocation figures. In our analysis the focus is on individual entities regulated by the Dutch supervisory authorities, because, in contrast to banking

[9] This data set does not include funeral-in-kind insurers and tiny mutual non-life insurers that are exempted from supervision.

[10] Because of the long time span of our data set, one could worry about structural breaks in the data. There have indeed been regulatory and accounting changes during our sample period. Fortunately, the timing of these breaks is known exactly. From 2006 and onwards, with the enactment of the Dutch Health Insurance Act, the former national health insurance funds enter our data set as non-life insurers, active in the health and accident line of business. For pension funds, there has been a regulatory regime change in 2007, as explained in Section 5.2.2. Hence, the period 1995-2005 constitutes a stable regime. We use this period to test the robustness of our full sample results.

[11] Though we exclude unit-linked *investments*, we account for unit-linked *activities* by including a dummy variable. Footnote 8 explains in detail how the line-of-business dummies are calculated.

[12] The raw data set contains 18,416 institution-year observations. We have excluded defined contribution (DC) pension funds (573 observations), as they would confound our analysis. In DC funds, the investment risk is typically borne by the pension fund participants and not by the pension fund itself. We have also excluded observations that have zero or negative total assets, negative equity allocation, equity allocation>100%, capital ratio>1 or capital ratio<-0.3. Note that our data source does not distinct between zeroes and missing values.

supervision, insurance supervision is non-consolidated. As most insurers are group affiliated, however, we control for group affiliation in our regression analyses.[13]

Table 5.2 summarizes the data. While life insurers are larger, on average, than pension funds, in terms of balance sheet size, the latter invest relatively more in equities.[14] On average, pension funds invest even more in equities than their surplus or buffer. This is generally not the case for life insurers and certainly not for non-life insurers. Non-life insurers have relatively large capital buffers, which apparently are not used to take on additional investment risk. As mentioned in the previous section, underwriting risk is their dominant risk category.

Table 5.2. Full sample summary statistics, monetary values in 2009 prices (1995-2009)

	Pension funds		Life insurers		Non-life insurers	
	mean	median	mean	median	mean	median
Equity holdings (in € millions)	372	15	365	12	28	1
Total investments (in € millions)	956	60	2204	286	145	15
Equity allocation (%)	28	28	14	8	16	9
Capital ratio (%)	10	12	15	9	47	47
Investments per participant (x 1000 €)	119	54				
Number of participants (x 1000)	26	1				
Share of active participants (%)	42	42				
Pension fund type						
Company	0.85	1				
Industry	0.14	0				
Professional group	0.01	0				
Reinsurance ratio (%)			6	1	24	15
Organizational form (1=mutual insurer)			0.11	0	0.45	0
Group (1=group affiliation)			0.73	1	0.65	1
Line-of-business dummies						
Traditional life			0.76	1		
Unit-linked			0.24	0		
Accident and health					0.29	0
Motor					0.08	0
Fire and other property risk					0.34	0
Marine, transport and aviation					0.05	0
Miscellaneous insurance					0.08	0
Institution-year observations	8,234		1,218		3,347	

Note: Figures are after selection. See Footnote 12 in this chapter for the selection criteria.

The average wealth per DB pension plan participant is about €120,000 (2009 prices), which is more than twice the median value. Boards of directors tend to have their own pension schemes and such schemes tilt average pension fund wealth up considerably. Notice that non-life insurers are relatively small investors compared to life insurers and pension funds. This is a direct result of the short-term nature of non-life insurance policies. The organizational-form

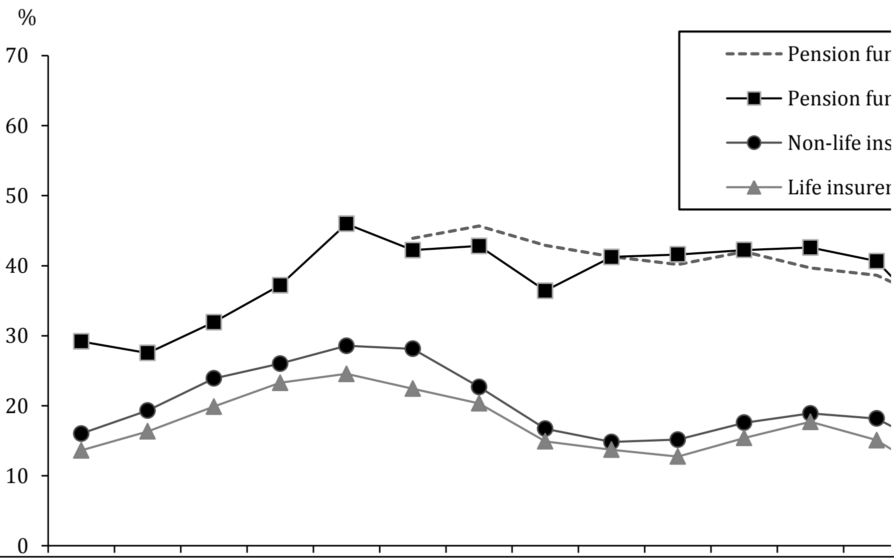
[13] As in Chapter 4, we take subsidiaries and not groups as the primary decision-making unit, because subsidiaries in the insurance industry tend to operate independently.

[14] World-wide, pension funds tend to have a higher equity allocation than life insurers (Broeders et al. 2011).

dummy shows that mutual insurers are most prevalent in the non-life industry. While the number of non-life insurers is almost three times that of life insurers, on average about 45% of the non-life insurers is mutual-owned.

Figure 5.1 plots the weighted-average equity allocation over our sample period. For pension funds, we do not only have actual equity allocations but also strategic asset allocations, as reported to DNB.[15] All institutional investors let their equity allocation increase during the late nineties, along with the technology bubble. When the bubble burst, however, investors' behavior diverged. While insurers let their equity exposure decrease, pension funds maintained equity allocation at a high level close to 40%, in line with their investment strategy. This points to negative feedback trading by pension funds. Indeed, Bikker et al. (2010) and De Haan and Kakes (2011) report over a shorter sample period – up to and including 2006 – systematic rebalancing behavior by Dutch pension funds.

Figure 5.1. Equity allocations by sector (weighted averages; 1995-2009)



Notes: Weighted-average equity allocations are total year-end equity holdings divided by total investments. For life insurers, unit-linked investments have been excluded.

Figure 5.1 shows that during the financial crisis of 2008-2009, pension fund were less successful in rebalancing their equity allocation. The weighted average equity allocation dropped by almost 10 percentage points in 2008. Interestingly, the strategic allocation was also

[15] For pension funds, the weighted-average equity allocation is considerably higher than the average equity allocation (as shown in Table 5.2), suggesting support for hypothesis *H5.1a* (risk taking increases in size).

adjusted downwards. Comparing life and non-life insurers, we see that the latter generally allocate a larger proportion of their assets to equities. We also see that life insurers were net seller of equities in 2009, as the equity allocation dropped while equity markets were up that year.

5.4 EMPIRICAL RESULTS

5.4.1 Equity allocations across investors

Because of the potential endogeneity of capital, we estimate Equation (5.1) using the two-stage-least-squares estimator (2SLS). The idea of 2SLS is to use a proxy for the (potentially) endogenous variable – here capital – that is uncorrelated with the disturbance term in the regression equation. This proxy is generated in the first stage of the 2SLS procedure, using instrumental variables. To obtain reliable results, the instrumental variables need to be both relevant (i.e. sufficiently correlated with the endogenous variable) and valid (i.e. uncorrelated with the disturbance term in the original equation). As instruments we use two lags of portfolio size, two lags of an insurance firms' profitability measure (i.e. profits after tax divided by total assets) and three lags of a pension funds' liability measure (i.e. the difference between contributions and pay-outs divided by total assets). Table 5.3 gives the 2SLS estimation results, where the last three rows indicate the instruments used are both relevant and valid.

Column I of Table 5.3 shows full sample results, while Column II presents outcomes for the shorter sample period 1995-2005. The shorter sample period is more stable in terms of regulatory and accounting regimes. As mentioned in Chapters 3 and 4, in 2006 the Health Insurance Act was implemented and the former national health insurance funds were privatized. In 2007, risk-based regulation of pension funds was introduced. Column III provides balanced-panel results for the 1995 to 2005 period. Differences between the balanced and unbalanced panel results may give an indication of the magnitude of survivorship bias. Of course, the varying set of observations may also have an effect on the parameter estimates.

Surprisingly, the capital ratio is insignificant for pension funds (α_1). This suggests that for pension funds as a group, neither risk-shifting nor risk-management incentives seem to dominate. This finding contrasts with research on US pension plans, where risk-management incentives have been found to dominate investment behavior (Rauh 2009). What could explain this difference? An important dissimilarity between the US and the Netherlands concerns the position of the sponsor in relation to the pension plan. While in the US, the sponsoring company fully bears the investment risk of its DB pension plan (until bankruptcy of course), in the Netherlands the investment risk is typically shared between the plan sponsor and the plan participants (Ponds and Van Riel 2009). Because of this difference in risk bearing, Dutch sponsors may well be relatively less worried about the risk of pension shortfalls, which could (partly) explain the insignificance of the capital ratio.

Table 5.3. 2SLS estimation results of equity-allocation model

Variable	Coefficient	1995-2009 unbalanced	1995-2005 unbalanced	1995-2005 balanced
Column		*I*	*II*	*III*
Constant	α_0	**28.72	**31.28	**32.78
Capital ratio (lagged)	α_1	4.41	-0.77	-13.46
Portfolio size (logarithm, lagged)	α_2	**2.47	**2.93	**2.60
D^I	α_3	**-13.69	**-13.55	**-11.43
D^I•*Capital ratio (lagged)*	α_4	**46.06	**48.73	**54.96
D^I•*Portfolio size (logarithm, lagged)*	α_5	**-1.44	**-1.60	**-1.87
D^I•*STOCK*	α_6	**5.20	**5.37	**5.43
$D^I \cdot LOB_{health}$	α_7	**-4.22	*-3.78	-3.10
$D^I \cdot LOB_{transport}$	α_8	**-9.70	**-7.05	-0.23
$D^I \cdot LOB_{fire}$	α_9	**-9.57	**-8.72	**-7.53
$D^I \cdot LOB_{miscellaneous}$	α_{10}	**-7.02	**-7.20	**-7.23
$D^I \cdot LOB_{traditional\ life}$	α_{11}	1.12	0.32	-0.28
$D^I \cdot LOB_{unit\text{-}linked\ life}$	α_{12}	**11.62	*10.03	*7.10
D^I•*Group affiliation*	α_{13}	0.11	0.09	-1.44
D^I•*Reinsurance ratio (lagged)*	α_{14}	0.33	*0.51	0.72
D^{PF}•*Share active participants*	α_{15}	**3.63	2.53	0.59
D^{PF}•*Assets per participant (logarithm, lagged)*	α_{16}	**1.48	**1.55	**2.03
D^{PF}•*Industry fund*	α_{17}	0.82	1.09	2.33
D^{PF}•*Professional group fund*	α_{18}	**-6.31	**-9.12	**-10.63
Number of observations		7,811	5,550	3,400
R^2, adjusted		0.22	0.25	0.31
First-stage F-test for *Capital ratio (lagged)*		38.43	19.80	15.89
First-stage F-test for D^I•*Capital ratio (lagged)*		56.68	49.40	25.03
Hansen test (p-value)		0.06	0.60	0.68

Notes: * and ** denote significance at, respectively, the 95% and 99% confidence level, calculated using Huber-White robust standard errors. The dependent in these regressions is the equity allocation in percentages. The capital ratio variables have been instrumented, using two lags of portfolio size, two lags of insurance firms' profitability and three lags a pension fund liability measure. The first-stage F-tests test the joint significance of these instruments for the two capital ratio variables. The Hansen test tests the joint null that the chosen instruments are valid. D^I is a dummy, which is 1 for insurers and 0 for pension funds. *LOB*. are the different line of business dummies, which take the value 1 for insurers that are predominantly active in the respective line, and 0 otherwise. Organizational form dummy *STOCK* is 1 for stock insurers and 0 otherwise. D^{PF} is a dummy, which is 1 for pension funds and 0 for insurers. Year dummy estimates are not shown. The year dummies are jointly significant (test results available on request). For expositional purposes, all non-dummy variables are in deviation from their sample means.

By contrast, for insurers the capital ratio is highly significant ($\alpha_1 + \alpha_4 > 0$), where the estimated parameters indicate risk-management incentives dominate. A positive relationship between capital and asset risk is also reported by Cummins and Sommer (1996) and Baranoff and Sager (2002) for property-liability and life insurers in the US, respectively. So, while we reject that investors with larger capital buffers take more equity risks in general (*H5.1a*), we find support for *H5.3* as insurers choose their asset allocations in a more risk-sensitive manner than pension funds.

Portfolio size is found to have a significant and positive effect on the investment risk taking of pension funds: when total investments increase by 1%, the equity allocation rises with 2 to 3 basis points (α_2). This result is in line with earlier estimates for Dutch pension funds (Bikker et al. 2010, 2012). For insurers, the relationship between portfolio size and risk taking is weaker (α_5 is significantly negative).[16] Nonetheless, the results are overall in line with hypothesis *H5.1b*. An explanation for the weaker relationship for insurers is that large pension funds may suffer more from overconfidence than large insurers. Although hard to prove scientifically, this explanation would be consistent with our other results. In line with *H5.2* we find that pension funds take significantly more investment risk than insurers, even after correcting for capital and size. The (weighted) sum of the insurer dummy coefficient (α_3) and the interaction coefficients $\alpha_6 - \alpha_{13}$ is significantly negative[17], which means that insurers have significantly lower equity allocations than pension funds.

Let us now turn to the hypothesis about underwriting risk and investment risk taking by insurers (*H5.1c*). The estimates of $\alpha_7 - \alpha_{10}$ show that health, transport, fire and miscellaneous insurers tend to take significantly less investment risk than motor insurers (reference group). Since motor insurers typically have relatively stable underwriting results (see Table 5.1), this finding supports *H5.1c*. Equity investments of unit-linked life insurers (to be sure: on their own account) are significantly larger than those of motor insurers, supporting *H5.1c*. Note the sign of the reinsurance ratio coefficient (α_{14}) also underpins *H5.1c*, yet this result is only significant for the shorter period unbalanced sample (Column II).

Hypotheses *H5.1d* and *H5.1e* concern pension funds. The estimates of α_{15} indicate that the share of active participants has a significantly positive effect for the 1995-2009 sample, but not for the 1995-2005 samples. One argument is that active participants largely determine the effectiveness of raising premium to restore the funding ratio. The life-cycle theory offers a second explanation, as it states that active (and younger) participants can take more risk – and, hence, should invest more in equities – than retired ones (Bikker et al. 2012). Once more, risk-bearing capacity is the main explanation: active participants can compensate negative returns

---

[16] The reported significant difference between insurers and pension funds with respect to the relationship between investment risk taking and portfolio size is also found when non-life insurers are excluded from the regression analysis. These specific regression results are not shown in Table 5.3 but are available upon request.

[17] The average equity allocation difference between pension funds and insurance firms is picked up by the insurance dummy. Since the insurance dummy is interacted with other dummy variables (i.e. *STOCK*, *LOB.* and *G)*, the equity-allocation difference between insurers and pension funds is a weighted sum of the estimates of coefficients α_3 and $\alpha_6 - \alpha_{13}$. Note that coefficients α_4, α_5 and α_{14} do not play a role in this respect, since the corresponding variables are in deviation from their respective sample means.

on investments by extending working hours or delaying retirement. Higher wealth per participant implies a significantly higher equity ratio (α_{16}), consistent with *H5.1e*.

Our last two hypotheses are about differences in risk taking between types of insurers and pension funds. The results in Table 5.3 show stock insurers indeed take significantly more investment risk than mutual insurers (α_6). Corrected for the relevant risk factors, stock insurers have an average equity allocation that is more than 5 percentage points higher than that of their mutual peers, confirming *H5.4*. Hypothesis *H5.5* states that professional-group funds take less investment risk than company funds and industry-wide funds, and this is supported by the estimates of α_{18}.

5.4.2 Equity allocations over time

Equity allocations are quite volatile over time. The question we try to answer here is to what extent this volatility is driven by investment behavior and whether this behavior can be related to our hypotheses on risk-bearing capacity and risk-taking incentives. Table 5.4 provides OLS and weighted-least-squares (WLS) estimation results of Equation (5.4). The WLS estimations weight the changes in equity ratios with real equity holdings, thereby providing insight into possible behavioral differences between small and large equity investors. The first four columns present the results for pension funds, while the last four columns present the results for insurance firms.

Table 5.4. Estimation results of equity-trading model (2000-2009)

	I	II	III	IV	V	VI	VII	VIII
	Pension funds				Insurance firms			
	OLS		WLS		OLS		WLS	
Equity returns	**0.58		*0.75		**0.81		0.97	
	(0.04)		(0.11)		(0.07)		(0.06)	
Positive returns		**0.55		0.71		**0.55		0.96
		(0.11)		(0.15)		(0.16)		(0.09)
Negative returns		**0.60		0.76		0.95		0.98
		(0.05)		(0.21)		(0.07)		(0.12)
R^2, adjusted	0.33	0.33	0.57	0.57	0.20	0.20	0.36	0.36
No. Obs.	4,451	4,451	4,451	4,451	2,151	2,151	2,151	2,151

Notes: * and ** denote significantly different from 1 at, respectively, the 5% and 1% level, calculated using Huber-White robust standard errors. The dependent variable in these regressions is the change in total equity holdings divided by year-end total investments. Only investors holding equities in their portfolio, either at the beginning or the end of the book year, are included in these regression analyses. The weighted-least-squares (WLS) estimation uses the absolute value of equity holdings (in 2009 prices) as weighting factor.

Several findings stand out. First, Columns I and II show that pension funds rebalance, on average, 40% of market price movements (in line with Bikker et al. 2010) and that rebalancing is more or less symmetric in terms of positive *versus* negative returns. The coefficients 0.55 and 0.60 in Column II represent free-floating behavior of around 60%, which is the complement of rebalancing. Second, insurers tend to rebalance about 45% of market price movements in bull markets, yet let their equity allocation fall in bear markets (column VI). Since negative feedback trading (or rebalancing) increases an investor's risk profile compared to a free-float strategy, these trading results provide further support to hypotheses *H2* (insurers take less investment risk) and *H3* (insurers are more risk sensitive in their investment behavior). Indeed, insurers moderate their risk taking when returns are negative and their risk-bearing capacity is down; such investment behavior is technically known as contingent immunization (Leibowitz and Weinberger 1982).

Third, comparing the OLS and WLS results for pension funds, we observe that large pension funds generally rebalance less than their smaller peers. This finding over 2000-2009 contrast with what has been found in earlier research on the investment behavior of Dutch pension funds by Bikker et al. (2010) and De Haan and Kakes (2011) over earlier periods. Though Bikker et al. (2010) use quarterly data, and we have annual observations, the frequency difference does not seem to drive our results. Indeed, we can replicate the results of Bikker et al. (2010) by taking approximately the same sample period (i.e. 2000-2006).[18] A possible explanation for the different behavior over our sample period is that large pension funds were unable to rebalance their sizeable equity losses of 2008 to the extent that they had done that before. Large pension funds' press statements dating from early 2009 are consistent with this interpretation (Kreijger 2009). This result is also relevant from a macro-perspective, since pension funds are typically perceived as providers of risk-bearing capital when it is most needed, yet did not perform this role at the height of the financial crisis. Notice the adjusted R^2 statistics of the WLS regressions are much larger than those of the OLS regressions, indicating that market returns are a better predictor of equity portfolio changes of large funds than those of small funds.

The fourth and final finding from Table 5.4 is that large insurance firms generally pursue a completely free-float investment strategy, whereas smaller insurers rebalance under favorable market conditions. Columns VII and VIII show that the estimated slope coefficients are not significantly different from 1, indicating that, on average, equity price movements feed roughly one-to-one into the equity allocation.

To investigate trading behavior across organizational form in the insurance industry, the first four columns of Table 5.5 show separate estimation results of Equation (5.4) for stock and mutual insurers. Note that the sample period is longer now: 1995-2009. The point estimates

[18] These results are not shown but available on request.

suggest stock insurers rebalance slightly more than mutual insurers, yet the differences are at best marginally significant. So, while we have found that the level of risk taking by stock insurers is larger than that of mutual insurers, investment behavior over time is not materially different.

The last four columns of Table 5.5 show estimation results for life and non-life insurers. The results in columns VI and VIII indicate that life insurers tend to rebalance considerably during downturns (44%), while non-life insurers seem to follow a strategy closer to free float. This greater investment risk taking by life insurers is consistent with hypothesis *H5.1c* as life insurers typically have less underwriting risk than non-life insurers. An alternative interpretation is that life insurers believe that rebalancing is optimal given their long investment horizon.

Table 5.5. OLS Estimation results of equity-trading model for insurance firms (1995-2009)

	I	II	III	IV	V	VI	VII	VIII
	Stock insurers		Mutual insurers		Life insurers		Non-life insurers	
Equity returns	*0.84		0.94		*0.77		0.91	
	(0.07)		(0.05)		(0.09)		(0.06)	
Positive returns		0.78		0.82		1.00		*0.73
		(0.14)		(0.16)		(0.09)		(0.12)
Negative returns		0.90		1.06		**0.56		1.07
		(0.09)		(0.04)		(0.14)		(0.07)
R^2, adjusted	0.19	0.19	0.40	0.40	0.19	0.20	0.25	0.26
No. Obs.	2,235	2,235	1,066	1,066	977	977	2,324	2,324

Notes: * and ** denote significantly different from 1 at, respectively, the 5% and 1% level, calculated using Huber-White robust standard errors. The dependent variable in these regressions is the change in total equity holdings divided by year-end total investments. Only insurers with positive equity holdings, either at the beginning or the end of the book year, are included in these regression analyses.

5.5 CONCLUSION

According to theory, institutional investors face both risk-management and risk-shifting incentives. When risk-management incentives are leading, lower risk-bearing capacity is accompanied by less investment risk taking. Conversely, if risk-shifting incentives are more prominent, worse financial conditions lead to more investment risk taking. This paper assesses the relevance of these conflicting incentives for Dutch pension funds and insurance firms over the period 1995-2009. Two measures of investment risk taking are used. The first measure is the allocation of equities in the total investment portfolio, which is intuitive and widely used. Our second measure concerns equity feedback trading. While there is a sizeable literature on feedback trading, the microprudential risk implications of such behavior have typically been passed over. Buying equities when equities are down (i.e. negative feedback trading) is a risky strategy, however, since the exposure to equities is raised when risk-bearing capacity is down.

In upward markets, positive feedback trading is relatively risky, that is compared to negative feedback trading or no feedback trading.

Using annual investments data, we conclude that risk-management incentives appear to dominate risk-shifting incentives in the Dutch insurance industry. Insurance firms with more risk-bearing capacity invest a larger share of their portfolio in equities, and vice versa. This result is in agreement with prior studies on US life and property-liability insurers. For Dutch pension funds, on the other hand, we do not find a significant relationship between the funding ratio and the asset allocation. Hence, neither risk-shifting nor risk-management incentives seem dominant. Also over time, pension funds are more risk tolerant. Pension funds rebalance, on average, about 40% of market price movements, in both bull and bear markets. Insurance firms also rebalance in bull markets but generally do not buy equities in bear markets to restore their equity allocation. This finding confirms that, in bear markets when risk-bearing capacity of insurers has been eroded, they are more risk averse. Insurance firms face direct and indirect costs of financial distress, while pension funds do not face bankruptcy risk under detrimental market conditions. Even so, unfavorable investment results do hit pension fund participants, just as they benefit from favorable outcomes. It is the responsibility of pension funds to adequately inform their participants about the consequences of investment risk taking on the safety of pension benefits.

Though risk-management incentives appear dominant among Dutch insurance firms, we also find evidence of risk-shifting behavior. In line with the ownership-structure hypothesis by Lee et al. (1997), stock insurance firms have more risky investments than their mutual peers. Interestingly, we find that professional group pension funds take significantly less investment risk than other types of pension funds. This result is in line with expectations, as the participants in professional group pension funds are typically self-employed, which excludes the possibility of risk shifting by the employer.

Chapter 6

CONCLUSION AND IMPLICATIONS

This thesis investigates the role and impact of incentives in the insurance industry. There is little disagreement among economists about the importance of incentives. As a result of the financial crisis, there is now also an increased understanding among policymakers that incentives are crucially important. Topical policy discussions about moral hazard in the banking system, the role of rating agencies in the subprime mortgage bubble and the effect of bonuses on risk taking are in essence all about getting the incentives right. When the incentives are properly aligned, (market) failures are much less likely to occur.

In the insurance industry, the relatively favorable crisis track record of insurance companies suggests well-aligned incentives. However, the profiteering policy affair in the Netherlands and the Payment Protection Insurance (PPI) affair in the UK suggest differently. These affairs points to short-term sales incentives that are not only damaging to policyholders but also pose a long-term threat to insurers. The large size of the Dutch insurance industry is often attributed – partly of course – to a high degree of risk aversion among the Dutch people. To what extent does this explanation make sense? And what has been the driving force behind the consolidation of the Dutch insurance industry over the past decades?

Inspired by the foregoing, five research questions have been formulated. Here we return to these questions and formulate answers based on the research findings in the preceding chapters. Section 6.2 discusses the research findings and provides key policy implications.

6.1 CONCLUSION

6.1.1 Advice incentives

In many countries, intermediaries are an important distribution channel for insurance products. While intermediaries are typically remunerated through sales commissions, there is increasing concern about conflicts of interests that arise from commission-based remuneration. Against this backdrop, the Netherlands and also Australia and the UK have recently prohibited sales commissions on certain financial products. While intermediaries have an incentive to provide suitable advice when commissions are banned, insurers also provide advice directly. Research Question 1 states:

RQ1: *To what extent does a ban on commissions incentivize insurers to bypass intermediaries and provide financial advice directly?*

Using a theoretical model, Chapter 2 concludes the extent to which insurers have an incentive to bypass intermediaries depends on the alertness of the policyholders. Customers that are alert about the incentives behind advice are willing to pay more for unbiased advice. Insurers therefore prefer to transact with such customers via an intermediary who is remunerated directly by the customers. Naïve customers, however, mistakenly believe the quality of direct advice is equivalent to that of fee-based intermediary advice. This misunderstanding of incentives makes it more profitable for financial institutions to deal with naïve customers directly. Thus, the more naïve policyholders are about advice incentives, the greater the incentive for insurers is to bypass the intermediaries and provide advice directly. This finding answers Research Question 1 and brings up Research Question 2, that is:

RQ2: *With multiple distribution channels, what is the expected impact of commission bans on the quality of financial advice?*

The main result of Chapter 2 is that there are theoretical reasons to suspect that commission bans won't improve the quality of financial advice. Inderst and Ottaviani (2012b) explain that there is an economic rationale for commission bans when customers do not adequately take into account the incentives behind advice. Chapter 2 shows, using an extended version of Inderst and Ottaviani's model, that such unawareness of advice incentives actually reduces the effectiveness of commission bans. The intuition behind this finding is straightforward. In the model's equilibrium, naïve customers receive direct advice, which is in terms of advice quality equivalent to intermediary advice influenced by product commissions, as analyzed by Inderst and Ottaviani.

6.1.2 Insurance demand incentives

The large size of the Dutch insurance industry is often casually ascribed to a great aversion to risk. To investigate the potential importance of risk aversion in explaining insurance demand, we study deductible choice in Dutch basic health insurance. The basic health insurance policy is offered by private health insurance companies and is mandatory. Adults have some degrees of freedom in choosing their insurance coverage. They can opt for a voluntary deductible, which comes on top of a mandatory deductible. In return for a higher deductible, policyholders receive a premium rebate. Interestingly, the large majority of Dutch adults choose a voluntary deductible of €0 and thus receive no premium rebate. Research Question 3 asks:

RQ3: *How important is risk aversion in explaining the limited appetite for nonzero voluntary deductibles in Dutch basic health insurance?*

Using population and survey data, Chapter 3 shows that risk aversion seems to dominate deductible choice. In the regression analyses, risk preferences are both statistically and economically more significant in explaining deductible choice than risk type. We find evidence that healthy individuals' monetary incentive to choose a high voluntary deductible is generally overshadowed by their risk aversion. Exemplarily is the choice behavior of men in their twenties. These young men tend to have particularly low out-of-pocket health expenses: in the year we study, 2008, seven out of ten had out-of-pocket expenses that did not exceed the mandatory deductible of €150. Even so, in the population, nine out of ten young males choose the lowest deductible possible.

6.1.3 Consolidation incentives

To gain insight into the consolidation of the Dutch insurance industry, we examine the restructuring of the Dutch non-life insurance industry. Over the period 1995-2005, this insurance industry underwent significant change. Over one fifth of non-life insurers seized to exist as standalone firms, and average premium income almost doubled in real terms. This consolidation did not lead to conglomeration of the industry. In fact, the market share of monoline firms active in one line of business only actually increased during these years. Over this same period, the market share of mutual insurers dropped substantially. Interestingly, these restructuring developments were preceded by substantial deregulation of the European insurance industry (with the exception of solvency regulation). As the objective of deregulation is to improve market efficiency through increased competition, Research Question 4 questions:

RQ4: *To what extent was the restructuring of the Dutch non-life sector driven by efficiency considerations?*

Using a translog cost function, Chapter 4 concludes cost-efficiency considerations have played a role in the restructuring of non-life industry, though not an entirely dominant one. For the majority of non-life insurers, increasing returns to scale are found. However, for the larger firms constant returns to scale or even diseconomies of scale are observed. This distinction between small and large insurers is relevant since the industry's consolidation has been primarily the result of acquisitions made by the largest insurers, not by the scale-inefficiently small ones. Had scale-efficiency considerations played a more dominant role, M&As at the lower end of the size distribution would have been more common. Another rationale for industry consolidation derives from X-efficiency differences between insurance firms. Inefficient insurers are a potential takeover target for insurers with more skilled management. Using a thick frontier approach, we observe large X-inefficiencies that have moderately decreased over time. Accordingly, this finding does seem to point to the importance of cost-efficiency considerations. The X-inefficiency estimates should be treated with due care, though, as firm-specific input prices lack in our translog cost function. A third finding is that more focused insurers are more cost-efficient, which clarifies the rise of monoline insurers. The deteriorating market share of mutual insurers does not seem to be the result of cost inefficiencies.

6.1.4 Investment risk-taking incentives

Insurers are rightly proud about their crisis track record, as it is generally better than that of banks and pension funds. Given the importance of investment returns for (life) insurers and the volatile financial markets during the recent crisis, insurers have presumably fared well thanks to cautious investment behavior. To investigate this more thoroughly, the investment behavior of Dutch insurance firms and DB pension funds is examined over the volatile investment period 1995-2009. Theory offers two main hypotheses about the relationship between risk-bearing capacity and risk taking. According to the risk-shifting hypothesis, asymmetry in the distribution of profits and losses provides an incentive to take more investment risk when risk-bearing capacity is low. Conversely, the risk-management hypothesis states that investor steer away from danger when risk-bearing capacity is low, to avoid direct and indirect bankruptcy costs. Research Question 5 is:

RQ5: *How do incentives affect the investment risk taking by insurance firms and pension funds respectively?*

Chapter 5 finds dominant risk-management incentives in the Dutch insurance industry. Insurance firms with more risk-bearing capacity are found to invest more in equities, and vice versa. By contrast, a significant relationship between the funding ratio and the equity allocation of Dutch pension funds is absent. Equity trading data further confirms this divergence between pension funds and insurance firms. While pension funds rebalance, on average, about 40% of market price movements, insurance firms generally refrain from buying equities in bear markets to restore their equity allocation. While rebalancing may have benefits from a macroprudential perspective, it also increases the riskiness of pension benefits compared to a free-float strategy.

6.2 DISCUSSION AND IMPLICATIONS

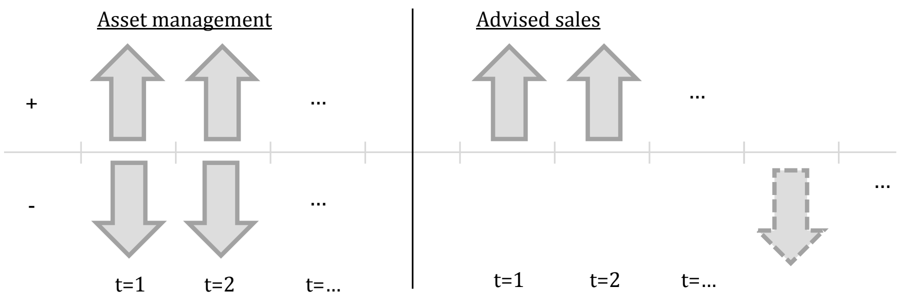
6.2.1 Discussion

Our findings on incentives in the insurance industry are mixed. With respect to asset management and insurance demand, our results point to well-aligned incentives. Indeed, we report evidence that insurance companies tend to manage risks, rather than shift risks, and that policyholders are willing to pay more than the expected value of an insurance policy because of risk aversion. In contrast, on financial advice we conclude that there is a genuine risk of undesirable outcomes, despite the recent ban on commissions. This is because a commission ban does not change insurers' incentive to advice the most profitable insurance policy, instead of the most suitable one.

It is puzzling that the incentives in asset management appear properly aligned, whereas the incentives behind financial advice remain skewed, despite policy intervention (i.e. banning commissions). The quality of financial advice has been shown to depend on customer alertness as well as on the costs of unsuitable advice. When both are low, the combination of sales and advisory tasks results into poor quality advice and welfare losses. In the Dutch profiteering policy affair, the costs of unsuitable advice have not been negligible, though. The redress costs are in the billions of euros (Verbond van Verzekeraars 2010) and the value impact of the reputation loss is perhaps even greater. The same holds for the PPI mis-selling scandal in the UK. According to FSA (2013), PPI was mis-sold on a large scale and lenders have had to pay £8 billion in redress already. So the costs of unsuitable advice and mis-selling can be sizeable.

This brings up the question of why the costs of unsuitable advice did not prevent unsuitable advice from happening in the first place. The theoretical model of Chapter 2 suggests that the expected benefits from unsuitable advice were simply greater than the costs. The timing and uncertainty of the payoffs seem also relevant. The costs of unsuitable advice typically take years to materialize. In fact, both in the profiteering policy affair and the PPI scandal, the total costs of unsuitable advice are still uncertain. By contrast, the 'costs' of risky investment behavior are quite clear ex-ante and can be incurred within the same day. This difference in potential payoffs of risk taking in asset management and advised sales is schematically depicted in Figure 6.1. As shown on the right-hand side, pursuing a risky sales strategy is strictly profitable in the short run, yet may backfire in the medium to long run. Such a payoff schedule encourages risk-shifting. Indeed, when the principals and agents of an insurance company benefit from high profits in the short run yet do not fully incur the losses in the longer term, there is an incentive to pursue a risky sales strategy. An asymmetry in payoffs may arise for agents/employees from sales commissions, and for principals/insurance owners from dividends and limited liability.

Figure 6.1 Possible payoffs from risk-taking, asset management vs. advised sales



Notes: This diagram is purely for illustrative purposes. Periods could refer to days, months, years, etc.

While deductible choice is clearly a different type of decision than product advice, in terms of payoffs there is some similarity with the advised-sales payoff schedule on the right-hand side of Figure 6.1. As discussed, the majority of Dutch adults currently opt for the lowest voluntary deductible in Dutch basic health insurance. These policyholders could opt for a higher deductible and therewith take more financial risk. A higher deductible brings short-term benefits in terms of lower monthly insurance premiums. That being so, a higher insurance deductible also exposes policyholders to potential future losses from higher out-of-pocket health expenditures. Since both the up- and the downside from a higher deductible are for the respective policyholder, there is no risk-shifting potential in this particular case. Our research has shown that risk aversion incentivizes most Dutch policyholders to prefer a higher insurance premium and more certainty to a lower insurance premium and more payoff volatility. Widespread risk aversion provides an explanation for the large size of the Dutch insurance industry. Fiscal incentives and institutional arrangements (private mandatory health insurance) are other explanatory factors.

The incentives behind the consolidation of the Dutch non-life insurance industry appear less well understood. While especially the largest non-life insurers have grown in size, returns to scale have been found to be zero or even negative at that end of the size distribution. For the Dutch life insurance industry, similar findings are reported by Bikker and Van Leuvesteijn (2008). In a widely-cited review article on consolidation in financial services, Berger, Demsetz, and Strahan (1999) conclude most types of consolidation have led to little or no cost-efficiency improvements on average. More dominant motives behind the growing size of financial institutions seem the drive for market power and profit efficiency.

6.2.2 Implications

The research findings have three main policy implications. First, on general level, policymakers should have more attention for incentives *ex-ante*, that is, before unfavorable outcomes occur. As a result of the recent financial crisis, there is now broad attentiveness for miss-aligned incentives in the financial industry. Moral hazard has become a commonly used word; bonuses have become highly controversial. This alertness is of course the result of the recent occurrence of unfavorable outcomes, e.g. mis-selling of financial products and failures of financial institutions. The risk exists that over time the lessons from the crisis start to fade. To mitigate this risk, policymakers should have a continuous attentiveness to incentives. To some extent, financial supervisors have taken this advice already on board. Indeed, supervisors increasingly take a forward-looking stance when assessing financial institutions. This is reflected in more regulatory attention to business models and strategy and incentive structures in the financial industry (Bank of England and Financial Services Authority 2011, DNB 2012). A close watch on incentives is especially rewarding in the insurance industry, given the long duration of many insurance products (in particular life annuities and disability insurance).

Second, in countries where commissions are banned, policymakers should pay special attention to direct advice. Closer attention to direct advice requires resources. The findings suggest that some of the required resources can be reallocated from the supervision of intermediaries. Not only are the incentives in the intermediary channel better aligned, because of the commission ban, the intermediary channel is also expected to attract the more alert customers. What exactly could conduct-of-business supervisors do to mitigate the risk of unsuitable direct advice? For one, more insight into direct sales processes could be achieved by deploying so-called mystery shoppers (FSA 2006b). These mystery shoppers act as potential customers and enable regulators to obtain a detailed understanding of the advice process early on. Lessons could also be learned from recent investigations in the UK by the FSA (2012). In light of the mis-selling of payment protection insurance in the UK, the FSA has documented numerous incentive schemes that increase the risk of unsuitable advice, and gives guidance on how conduct-of-business risks to customers could be addressed. Chapter 2 also highlights the possibility of improving the quality of advice by raising the regulatory penalty for unsuitable advice. Finally, conduct-of-business supervisors could initiate research on customers' awareness of advice incentives, and especially the conflicts of interests inherent in direct advice. Dependent on the results, product advice by financial institutions may need to come with a warning that the advice given is not independent, to increase customer alertness. Policymakers should be aware, however, that the disclosure of conflicts of interests may act as a moral-license for self-interested behavior (Cain, Loewenstein, and Moore 2005).

Third, prudential supervisors should be wary for a growing importance of risk-shifting incentives in the insurance industry. Our finding of strong risk-management incentives in the Dutch insurance industry is of course a welcome finding from a supervisory perspective. We also find evidence of risk-shifting incentives, however. For example, stock-owned insurers are found to take more investment risk than mutual insurers, presumably because the benefits of risk-shifting are greater. Now that competition has increased and economic profitability has decreased, especially in the Dutch life insurance industry (Bikker 2012), risk-shifting incentives are expected to increase. Indeed, there is a sizeable literature that suggests a negative relationship between risk taking and franchise value, where the latter is the value of the economic profits that a firm is expected to earn going concern. Intuitively, insurers with more franchise value have more to lose. Given that franchise value is down in the Dutch life insurance industry, risk-taking incentives have increased. There are signs that increased risk taking is happening already, as DNB (2011) warns for risky pricing strategies among Dutch life insurers.

The results of this thesis also have implications for future research. The finding of significant risk aversion in the Dutch population deserves further investigation. It is consistent with the popular impression that the Dutch people are rather risk averse and consequently highly insured. However, this result has been obtained in a specific context, i.e. deductible choice in a mandatory health insurance market, so care should be taken when generalizing. More empirical research on insurance demand, preferably in other contexts, would be welcome. The same holds for the consolidation of the Dutch insurance industry, which is not well understood. As returns to scale are probably low or even negative for the largest insurers, why are these insurers growing ever larger?

The theoretical chapter on financial advice provides clear testable hypotheses that could be tested in subsequent empirical research. First, the model implies that insurers prefer multi-channel distribution strategies when commissions are banned. Now that commissions are actually banned in the Netherlands, we expect that direct advice and multi-channel strategies become more common practice. Second, from the model we expect customers to segment themselves: alert customers opt for intermediary advice whereas naïve customers choose direct advice. This hypothesis can also be tested, for example via survey research. Accordingly, research can inform policymakers on the impact of commission bans and may help to prevent undesirable outcomes in the years to come.

Chapter 7

SAMENVATTING
(SUMMARY IN DUTCH)

Wat begon als een schijnbaar triviale gebeurtenis op de Amerikaanse hypotheekmarkt in 2007, is nadien uitgegroeid tot een wereldwijde financiële crisis met ernstige economische gevolgen. Verzekeraars hebben deze crisis relatief goed doorstaan. Terwijl de meerderheid van de banken staatssteun heeft ontvangen - expliciet dan wel impliciet - en veel pensioenfondsen in onderdekking zijn geraakt, zijn faillissementen onder verzekeraars relatief schaars gebleven. Hieruit zou men kunnen concluderen dat verzekeraars in het algemeen worden geprikkeld om voorzichtig te opereren en het klantbelang centraal te stellen.

De 'woekerpolisaffaire' schept een ander beeld. Met name eind jaren '90 van de vorige eeuw zijn miljoenen beleggingsverzekeringen verkocht in Nederland. Veel van deze verzekeringen waren buitenproportioneel kostbaar voor polishouders, maar bijzonder winstgevend voor verzekeraars en tussenpersonen, althans op de middellange termijn. Boot (1995) kaartte de problematiek al vroegtijdig aan, maar pas sinds 2006 bestaat brede maatschappelijke kritiek op beleggingsverzekeringen. Door de frequentie en hevigheid van de berichtgeving is de beleggingsverzekeringsmarkt niettemin in korte tijd opgedroogd. Tegenwoordig hangen de in het verleden zo winstgevende beleggingsverzekeringen de Nederlandse levensverzekeraars als een molensteen om de nek. Miljarden euro's zijn al aan kostenvergoedingen betaald en de claims stapelen zich op.

Het voorgaande onderstreept het belang van prikkels in de (Nederlandse) verzekeringssector. Het suggereert ook dat prikkels in de beleggingssfeer anders uitwerken dan bij de polisverkoop. Terwijl verzekeraars kennelijk zijn geprikkeld tot prudent beleggingsbeleid, gericht op de lange termijn, duidt de woekerpolisaffaire op schadelijke kortetermijnprikkels in de adviessfeer. Beleggingsverzekeringen zijn veelal door tussenpersonen geadviseerd, die daartoe werden geprikkeld via provisies. Beleidsmakers zijn in toenemende mate bezorgd dat indirecte betaling voor advies via provisies leidt tot belangenverstrengeling en daarmee tot slechte advieskwaliteit. Daarom is onder andere in Nederland en het Verenigd Koninkrijk voor bepaalde producten een provisieverbod ingevoerd. In Nederland geldt dit verbod per 1 januari 2013 voor alle verzekeringen, met uitzondering van 'eenvoudige' schadeverzekeringen. Het doel van een provisieverbod is tot een verbetering van de advieskwaliteit te komen.

Hoofdstuk 2 onderzoekt de kwaliteit van financieel advies wanneer provisies verboden zijn. Met behulp van een micro-economisch model laat hoofdstuk 2 zien dat een provisieverbod het aantrekkelijk maakt voor verzekeraars om gelijktijdig twee advieskanalen te bestieren: een intermediair kanaal en een direct kanaal. Het intermediaire kanaal is aantrekkelijk voor alerte klanten die de prikkels achter advies volledig doorgronden. Omdat deze klanten de intermediair rechtstreeks belonen voor advies, is de advieskwaliteit optimaal. Klanten met minder oog voor adviesprikkels geven echter de voorkeur aan direct advies.

De kernbevinding van hoofdstuk 2 is dat duidelijke vraagtekens zijn te plaatsen bij de effectiviteit van een provisieverbod. Inderst en Ottaviani (2012b) concluderen dat een provisieverbod wenselijk is wanneer klanten onvoldoende rekening houden met de adviesprikkels van tussenpersonen. Hoofdstuk 2 toont aan dat een dergelijke naïviteit onder klanten de effectiviteit van een provisieverbod juist sterk vermindert. In het model verbetert de advieskwaliteit zelfs helemaal niet, omdat direct advies in termen van advieskwaliteit gelijk is aan intermediair advies gestuurd door provisies.

Het provisieverbod is een actueel beleidsthema in Nederland; de omvang van de financiële sector is dat ook. Concentreren we ons op de verzekeringssector, dan is deze inderdaad als groot te kwalificeren, zowel in absolute zin als in verhouding tot de Nederlandse economie. De grote omvang wordt vaak terloops toegeschreven aan de sterke risicoaversie van de Nederlandse bevolking. Om de validiteit van deze redenering te kunnen duiden, onderzoekt hoofdstuk 3 de eigen risicokeuze in de Nederlandse basisverzekering. De basisverzekering wordt aangeboden door particuliere zorgverzekeraars en is verplicht voor alle inwoners. Volwassenen kunnen kiezen voor een vrijwillig eigen risico, dat bovenop het verplicht eigen risico komt. In ruil voor een hoger eigen risico, ontvangen verzekerden een premiekorting. Niettemin kiest de overgrote meerderheid van de Nederlanders voor een vrijwillig eigen risico van €0 en ontvangt dus geen premiekorting. De vraag rijst in hoeverre dit keuzegedrag wordt bepaald door risicoaversie.

Op basis van populatie- en steekproefgegevens over 2008, concludeert hoofdstuk 3 dat risicoaversie de eigen risicokeuze domineert. Regressieanalyses laten zien dat risicopreferenties een belangrijkere determinant van de eigen risicokeuze zijn dan geldprikkels, zowel in statistische als in economische zin. Exemplarisch is het keuzegedrag van jonge mannen. Terwijl zeven op de tien mannelijke twintigers zorgkosten hebben die het verplicht eigen risico niet overstijgen (€150 in 2008), kiest minder dan één op de tien voor een positief vrijwillig eigen risico. Merk op dat dit een substantiële groep in de Nederlandse samenleving betreft: bijna één miljoen mannen zit in die leeftijdscategorie. De overgrote meerderheid hiervan laat dus een verwacht financieel voordeel liggen. Deze bevindingen stroken met de communis opinio dat Nederlanders tamelijk risicoavers zijn en daarom flink verzekerd. Overigens zijn fiscale prikkels en institutionele aspecten (privaat zorgverzekeringsstelsel) ook belangrijke verklaringen voor de omvang van de Nederlandse verzekeringssector.

Hoewel de Nederlandse verzekeringssector omvangrijk is, groeit deze de laatste jaren weinig, en is in deelsegmenten (met name individueel leven) zelfs sprake van krimp. Door consolidatie zijn de verzekeraars wel steeds groter geworden. Om inzicht te krijgen in de prikkels achter deze consolidatieslag, onderzoekt hoofdstuk 4 de herstructurering van de Nederlandse schadeverzekeringssector. Over de periode 1995-2005 nam het aantal schademaatschappijen met 20% af, en verdubbelde de gemiddelde premieomzet in reële termen. Opmerkelijk genoeg leidde deze consolidatie van de sector niet tot minder focus: het marktaandeel van maatschappijen actief in één schadebranche (monolines) nam toe, niet af. Over dezelfde periode zagen onderlinge maatschappijen hun marktaandeel afnemen, ten gunste van verzekeraars met een aandeelhoudersstructuur. De timing van de herstructurering suggereert dat efficiëntieoverwegingen een belangrijke rol hebben gespeeld. In de jaren daarvoor had in Europa namelijk een dereguleringsslag plaatsgevonden, die tot doel had de concurrentie en keuzemogelijkheden in de nationale verzekeringsmarkten te vergroten. De invoering van een Europees Paspoort in 1994 maakte het verzekeraars mogelijk om op basis van een nationale vergunning in heel Europa actief te zijn. Verder werd in diverse Europese landen pas in de jaren negentig gestopt met product- en prijsregulering; dit geldt niet voor Nederland, maar bijvoorbeeld wel voor Duitsland.

Hoofdstuk 4 concludeert dat kostenefficiëntie-overwegingen een rol hebben gespeeld in de herstructurering van de schadeverzekeringssector, maar vermoedelijk niet dominant zijn geweest. Dit laatste blijkt onder andere uit de geschatte schaaleffecten. Terwijl de meeste schadeverzekeraars schaalvoordelen onbenut laten, geldt dit niet voor de grootste maatschappijen. De grootste schademaatschappijen lijken de optimale schaalgrootte te hebben of zijn wellicht zelfs te groot vanuit kostenefficiëntie-perspectief, wat moeilijk te rijmen is met de leidende rol van de grote verzekeraars in de waargenomen consolidatieslag. Een andere mogelijke grond voor consolidatie zijn operationele kosteninefficiënties (X-inefficiënties). Inefficiënte verzekeraars zijn een potentieel overnamedoelwit voor beter geleide verzekeraars. We constateren grote X-inefficiënties, die enigszins afnemen over de onderzoeksperiode. Deze bevinding lijkt het belang van efficiëntie-overwegingen te bevestigen, hoewel de X-efficiëntieschattingen wel met relatief veel onzekerheid zijn omgeven. Een derde bevinding is dat meer gefocuste verzekeraars kostenefficiënter zijn, wat de opkomst van monoline verzekeraars kan verklaren. Tot slot lijkt het afnemende marktaandeel van onderlinge verzekeraars niet het gevolg van kosteninefficiënties.

Zoals al aangegeven, hebben verzekeraars de financiële crisis beter doorstaan dan banken en pensioenfondsen. Tegen deze achtergrond onderzoekt hoofdstuk 5 de beleggingsprikkels van Nederlandse pensioenfondsen en verzekeraars over de volatiele periode 1995-2009. Gezamenlijk beheerden deze institutionele beleggers in 2009 ruim één biljoen euro aan beleggingen. Volgens de financieel-economische theorie hebben beleggers zowel risicobeheersingsprikkels als risicoverschuivingsprikkels. De prikkel om risico's te beheersen vloeit primair voort uit directe en indirecte faillissementskosten; de prikkel om risico's te verschuiven kan ontstaan bij een asymmetrie in de verdeling van winsten en verliezen. Als gevolg van beperkte aansprakelijkheid, is van zo'n asymmetrie bijvoorbeeld sprake binnen ondernemingen met een aandeelhoudersstructuur. Maar ook bij pensioenfondsen kan een prikkel bestaan om meer risico te nemen, bijvoorbeeld bij de werkgever omdat deze zijn bijdrage aan het fonds heeft begrensd. Om het belang van beide type prikkels vast te stellen, wordt de aandelenallocatie in de beleggingsportefeuille gebruikt als maatstaf voor het beleggingsrisico. De aandelenbeleggingen worden zowel in de cross-sectie als over de tijd onderzocht.

De hoofdconclusie van hoofdstuk 5 is dat in de verzekeringsindustrie de risicobeheersingsprikkels sterker zijn dan de risicoverschuivingsprikkels. Verzekeraars met meer solvabiliteit beleggen meer in aandelen, en vice versa. Voor pensioenfondsen wordt daarentegen geen significant verband gevonden tussen de dekkingsgraad en de aandelenallocatie. Het beleggingsgedrag over de tijd laat eenzelfde beeld zien. Pensioenfondsen kopen aandelen bij wanneer de beurskoersen dalen; een dergelijke beleggingsstrategie wordt in jargon herbalancering genoemd, omdat de aandelenallocatie weer richting het oude nivea.u wordt gebracht. Herbalancering is een riskante strategie in vergelijking tot de 'free-float' strategie die de meeste verzekeraars hanteren. Pensioenfondsen onderbouwen het gebruik van een herbalanceringsstrategie veelal – impliciet dan wel expliciet – door op 'mean reversion' in aandelenprijzen te wijzen. Als van mean reversion sprake is, tenderen aandelenprijzen naar een langetermijntrend, wat het bijkopen in een baisse kan motiveren. Recent onderzoek laat echter zien dat de mate van mean reversion in aandelenprijzen vermoedelijk klein is en bovendien onzeker.

Samenvattend, duiden de resultaten er op dat bij vermogensbeheer en de vraag naar verzekeringen de prikkels vanuit welvaartsperspectief goed liggen. In het onderzoek naar de consolidatie in de Nederlandse schadesector zijn de belangrijkste prikkels mogelijk niet aan het licht gekomen. Kostenoverwegingen lijken in ieder geval niet dominant te zijn geweest. Bij financieel advies blijkt het risico op onwenselijke prikkels significant. Het provisieverbod heeft weliswaar de prikkels voor tussenpersonen verbeterd, maar bij direct advies bestaat nog altijd de drijfveer om de meest winstgevende polissen te adviseren, in plaats van de meest geschikte.

Uit de bevindingen volgen ten minste drie beleidsimplicaties. Ten eerste moeten beleidsmakers *ex-ante* meer aandacht hebben voor prikkels, dat wil zeggen voordat problemen zich voordoen. Als gevolg van de financiële crisis, bestaat momenteel brede aandacht voor prikkels in de financiële sector. Moral hazard is onderdeel geworden van het normale spraakgebruik; bonussen zijn tegenwoordig zeer omstreden. De huidige alertheid is natuurlijk het gevolg van recente ongelukken. Het gevaar bestaat dat na verloop van tijd de lessen uit de crisis beginnen te vervagen, en daarmee de aandacht voor prikkelstructuren. Om dit risico te beperken, zullen beleidsmakers een continue aandacht moeten hebben voor prikkels. Financieel toezichthouders hebben dit advies in zekere zin al omarmd. Toezichthouders kiezen in toenemende mate een vooruitblikkend perspectief bij de beoordeling van financiële instellingen.

Dit komt onder andere tot uiting in meer aandacht voor bedrijfsmodellen en strategie en prikkelstructuren (Bank of England en de Financial Services Authority 2011, DNB 2012). Een goed begrip van prikkels is in het bijzonder de moeite waard in de verzekeringssector. Door de lange looptijd van veel verzekeringen (bijv. lijfrenten en uitvaartverzekeringen), kunnen misstanden namelijk lang onopgemerkt blijven. Met de juiste prikkels wordt de kans op misstanden structureel verkleind.

Ten tweede is in landen met een provisieverbod, zoals Nederland en het Verenigd Koninkrijk, extra toezichtaandacht gewenst voor direct advies door financiële instellingen. De hiervoor benodigde capaciteit kan mogelijk (deels) worden uitgespaard bij het gedragstoezicht op tussenpersonen. Niet alleen liggen – dankzij het provisieverbod – de adviesprikkels in het tussenpersoonkanaal beter, ook laat ons model zien dat het intermediaire kanaal vermoedelijk de alertere klanten trekt. Wat kunnen gedragstoezichthouders concreet doen om de risico's in het directe advieskanaal te verminderen? Een beleidsoptie is de inzet van zogenaamde 'mystery shoppers' (FSA 2006b). Mystery shoppers spelen potentiële klanten van financiële instellingen en stellen toezichthouders in staat om vroegtijdig een gedetailleerd inzicht te verkrijgen in het adviesproces en de advieskwaliteit. Daarnaast kunnen lessen worden getrokken uit recente onderzoeken in het Verenigd Koninkrijk door de FSA (2012). In het licht van de misleidende verkoop van betalingsbeschermers in het Verenigd Koninkrijk, heeft de FSA diverse risicovolle adviesprikkels gedocumenteerd. Ook geeft de FSA aan hoe financiële instellingen deze prikkels kunnen veranderen opdat de advieskwaliteit verbetert. Ten slotte kunnen de kosten van benedenmaats advies ook meer op de voorgrond worden gebracht via toezichtboetes. Uit hoofdstuk 2 volgt dat de advieskwaliteit daarmee verbetert.

Ten derde moeten prudentieel toezichthouders alert zijn op een toenemend belang van risicoverschuivingsprikkels in de verzekeringssector. Onze bevinding van sterke risicobeheersingsprikkels in de Nederlandse verzekeringssector is natuurlijk een welkome bevinding, niet in de laatste plaats vanuit toezichtperspectief. Verzekeraars met een goede risicobeheersing vragen immers minder toezichtaandacht. Door toegenomen concurrentie en afgenomen economische winstgevendheid, met name in de Nederlandse levensverzekeringssector (Bikker 2012), zijn de risicoverschuivingsprikkels in de verzekeringssector echter toegenomen. De aandeelhouders van verzekeraars hebben tegenwoordig minder te verliezen dan vroeger, toen de winsten torenhoog waren. DNB's (2011) waarschuwing voor te scherpe prijsstelling in de Nederlandse levensverzekeringssector duidt op waakzaamheid van de toezichthouder voor te risicovol gedrag.

REFERENCES

Alessie, R.J.M., S. Hochguertal, and A. van Soest, 2004, "Ownership of Stocks and Mutual Funds: A Panel Data Analysis," *Review of Economics and Statistics* 86 (3), 783-796.

Alter, G., and J.C. Riley, 1986, "How to Bet on Lives: A Guide to Life Contingent Contracts in Early Modern Europe," *Research in Economic History* 10, 1-53.

Anagol, S., S. Cole, and S. Sarkar , 2012, "Understanding the Incentives of Commissions Motivated Agents: Theory and Evidence from the Indian Life Insurance Market," Harvard Business School Working Paper 12-055.

Arrow, K.J., 1971, *Essays in the theory of risk-bearing*. Markham Publishing Company, Chicago.

Balvers, R., Y. Wu, and E. Gilliland, 2000, "Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies," *Journal of Finance* 55 (2), 745–772.

Bank of England and Financial Services Authority, 2011, "The Bank of England, Prudential Regulation Authority: Our Approach to Banking Supervision".

Baranoff, E.G., and T.W. Sager, 2002, "The Relations among Asset Risk, Product Risk, and Capital in the Life Insurance Industry," *Journal of Banking and Finance* 26 (6), 1181–1197.

Baranoff, E.G., and T.W. Sager, 2003, "The Relations among Organizational and Distributional Forms and Capital and Asset Risk Structures in the Life Insurance Industry," *Journal of Risk and Insurance* 70 (3), 375-400.

Barberis, N., M. Huang, and R.H. Thaler, 2006, "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case of Narrow Framing," *American Economic Review* 96 (4), 1069-1090.

Bar-Isaac, H., G. Caruana, and V. Cuñat, 2010, "Information Gathering and Marketing," *Journal of Economics and Management Strategy* 19 (2), 375-401.

Barsky, R.B., Juster, F.T., Kimball, M.S., and M.D. Shapiro, 1997, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics* 112 (2), 537-579.

Berger, A.N., J.D. Cummins, and M.A. Weiss, 1997, "The Coexistence of Multiple Distribution Systems for Financial Services: The Case of Property-Liability Insurance," *Journal of Business* 70 (4), 515-546.

Berger, A.N., J.D. Cummins, M.A. Weiss, and H. Zi, 2000, "Conglomeration vs. Strategic Focus: Evidence from the Insurance Industry," *Journal of Financial Intermediation* 9 (4), 323-362

Berger, A.N., and D.B. Humphrey, 1991, "The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking," *Journal of Monetary Economics* 28 (1), 117-148.

Berger, A.N., and D.B. Humphrey, 1997, "Efficiency of Financial Institutions: International Survey and Directions for Future Research," *European Journal of Operations Research* 98 (2), 175–212.

Bertaut, C.C., 1998, "Stockholding Behavior of U.S. Households: Evidence from the 1983-1989 Survey of Consumer Finances," *Review of Economics and Statistics* 80 (2), 263-275.

Bester, H., 1998, "Quality Uncertainty mitigates Product Differentiation," *RAND Journal of Economics* 29 (4), 828-844.

Bikker, J.A., 2012, "Performance of the Life Insurance Industry under Pressure: Efficiency, Competition and Consolidation," Working Paper No. 357. De Nederlandsche Bank, Amsterdam.

Bikker, J.A., D.W.G.A. Broeders, and J. de Dreu, 2010, "Stock Market Performance and Pension Fund Investment Policy: Rebalancing, Free Float, or Market Timing?," *International Journal of Central Banking* 6 (2), 53–79.

Bikker, J.A., D.W.G.A Broeders, D.A. Hollanders, and E.H.M. Ponds, 2012, "Pension Funds' Asset Allocation and Participant Age: A Test of the Life-Cycle Model," *Journal of Risk and Insurance* 79 (3), 595-618.

Bikker, J.A., and J.K. Gorter, 2011, "Restructuring of the Dutch Non-Life Insurance Industry: Consolidation, Organizational Form and Focus," *Journal of Risk and Insurance* 78 (1), 163-184.

Bikker, J.A., and M. van Leuvensteijn, 2008, "Competition and Efficiency in the Dutch Life Insurance Industry," *Applied Economics* 40 (16), 2063–2084.

Bikker, J,A., and P.J.G. Vlaar, 2007, "Conditional Indexation of Defined Benefit Pension Plans," *Geneva Papers on Risk and Insurance – Issues and Practice* 32 (4), 494–515.

Black, F., 1988, "An equilibrium model of the crash," in *NBER Macroeconomics Annual 1988*, ed. by S. Fischer. MIT Press, Boston.

Bolton, P., X. Freixas, and J. Shapiro, 2007, "Conflicts of Interest, Information Provision, and Competition in Banking," *Journal of Financial Economics* 85 (2), 297-330.

Boot, A.W.A., 1995, "Koopsommen en Premie-Stortingen Een Goudmijn: Maar voor Wie?, " available at: www.accf.nl/uploads/1995,%20onderzoek%20over%20woekerpolissen.pdf (accessed 8 February 2013).

Broeders, D.W.G.A., A. Chen, and B. Koos, 2011, "A Utility-Based Comparison of Pension Funds and Life Insurance Companies under Regulatory Constraints," *Insurance: Mathematics and Economics* 49 (1), 1-10.

Brunnermeier, M.K., 2009, "Deciphering the Liquidity and Credit Crunch," *Journal of Economic Perspectives* 23 (1), 77–100.

Buchmueller, T.J., Fiebig, D., Jones, G., and E. Savage, 2008, "Advantageous Selection in Private Health Insurance: The Case of Australia," Working Paper 2008/2. University of Technology, Sydney.

Cain, D.M., G. Loewenstein, and D.A. Moore, 2005, "The Dirt on Coming Clean: Perverse Effect of Disclosing Conflicts of Interests," *Journal of Legal Studies* 34 (1), 1-25.

Carlin, B.I., 2009, "Strategic Price Complexity in Retail Financial Markets," *Journal of Financial Economics* 91 (3), 278-287.

Carroll, C.D., 1994, "How Does Future Income Affect Current Consumption?," *Quarterly Journal of Economics* 109 (1), 111-147.

Chater, N., S. Huck, and R. Inderst, 2010, "Consumer Decision-Making in Retail Investment Services: A Behavioral Economics Perspective," Report to the European Commission/SANCO.

Chetty, R., and A. Szeidl**,** 2007, "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics* 122 (2), 831-877.

Chiang, L.C., and J.A. Krosnick, 2010, "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment," *Public Opinion Quarterly* 74 (1), 154-167.

Chiappori, P-A., and B. Salanié, 2008, "Modeling Competition and Market Equilibrium in Insurance: Empirical Issues," *American Economic Review* 8 (2), 146-150.

Cho, D., 1988, "Some Evidence of Scale Economies in Workers' Compensation Insurance," *The Journal of Risk and Insurance* 55 (2), 324-330.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, 1973, "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics* 55 (1), 28–45.

Cicchetti, C.J., and J.A. Dubin, 1994, "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy* 102 (1), 169-186.

Cohen, A., 2005, "Asymmetric Information and Learning in the Automobile Insurance Market," *Review of Economics and Statistics* 87 (2), 197-207.

Cohen, A., and L. Einav, 2007, "Estimating Risk Preferences from Deductible Choice," *American Economic Review* 97 (3), 745-788.

Cohen, A., and P. Siegelmann, 2010, "Testing for Adverse Selection in Insurance Markets," *Journal of Risk and Insurance* 77 (1), 39-84.

Cohn, R.A., W.G. Lewellen, R.C. Lease, and G.G. Schlarbaum, 1975, "Individual Investor Risk Aversion and Investment Portfolio Composition," *Journal of Finance* 30 (2), 605–620.

Commonwealth Treasury, 2012, "Future of Financial Advice," available at: http://futureofadvice.treasury.gov.au (accessed 16 May 2012).

Cummins, J.D., and P.M. Danzon, 1997, "Price, Financial Quality and Capital Flows in Insurance Markets," *Journal of Financial Intermediation* 6 (1), 3–38.

Cummins, J.D., and G. Nini, 2002, “Optimal Capital Utilization by Financial Firms: Evidence from the Property-Liability Insurance Industry,” *Journal of Financial Services Research* 21 (1), 15-53.

Cummins, J.D., and M. Rubio-Misas, 2006, “Deregulation, Consolidation and Efficiency: Evidence from the Spanish Insurance Industry,” *Journal of Money, Credit and Banking* 38 (2), 323-355.

Cummins J.D., M. Rubio-Misas, and H. Zi, 2004, “The Effect of Organizational Structure on Efficiency: Evidence from the Spanish Insurance Industry,” *Journal of Banking and Finance* 28 (12), 3113–3150.

Cummins, J.D., and D.W. Sommer, 1996, “Capital and Risk in Property-Liability Insurance Markets,” *Journal of Banking and Finance* 20 (6), 1069–1092.

Cummins, J.D., and M.A. Weiss, 1993, “Measuring Cost Efficiency in the Property-Liability Insurance Industry,” *Journal of Banking and Finance* 17 (2-3), 463-481.

Cummins, J.D., and M.A. Weiss, 2000, “Analyzing Firm Performance in the Insurance Industry Using Frontier Efficiency and Productivity Methods,” in *Handbook of Insurance*, ed. by G. Dionne. Kluwer Academic Publishers, Boston.

Cutler, D.M., Finkelstein, A., and K. McGarry, 2008, “Preference Heterogeneity in Insurance Markets,” *American Economic Review* 98 (2), 157-162.

Cutler, D. M., and R.J. Zeckhauser, 2000, “The Anatomy of Health Insurance,” in *Handbook of Health Economics*, ed. by A.J. Culyer, and J.P. Newhouse. Elsevier Science, Amsterdam.

Dahlby, B., 1983, “Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance,” *Journal of Public Economics* 20 (1), 121-130.

Dahlby, B., 1992, Testing for Asymmetric Information in Canadian Automobile Insurance, in *Contributions to Insurance Economics*, ed. by G. Dionne. Kluwer Academic Publishers, Boston.

Davis, E.P., 2002, “Prudent Person Rules or Quantitative Restrictions? The Regulation of Long Term Institutional Investors’ Portfolios,” *Journal of Pensions Economics and Finance* 1 (2), 157-191.

De Nederlandsche Bank (DNB), 2011, “Winstgevendheid Levensverzekeringssector onder Druk”.

De Nederlandsche Bank (DNB), 2012, “DNB Supervisory Strategy 2010-2014 and Themes 2010”.

Diecidue, E., and P.P. Wakker, 2001, “On the Intuition of Rank-Dependent Utility,” *Journal of Risk and Uncertainty* 23 (3), 281-298.

Doherty, N.A., and G. Dionne, 1993, “Insurance with Undiversifiable Risk: Contract Structure and Organizational Forms of Insurance Firms,” *Journal of Risk and Uncertainty* 6 (2), 187-203.

Doiron, D., G. Jones, and E. Savage, 2008, “Healthy, Wealthy and Insured? The Role of Self Assessed Health in the Demand for Private Health Insurance,” *Health Economics* 17 (3), 317-334.

Duan, N., 1983, “Smearing Estimate: a Nonparametric Retransformation Method,” *Journal of the American Statistical Association* 78 (383), 605-610.

Eling, M. and M. Lunen, 2010, “Frontier Efficiency Methodologies to Measure Performance in the Insurance Industry: Overview, Systematization, and Recent Developments,” *Geneva Papers on Risk and Insurance Issues and Practice* 35 (2), 217-265.

Ellison, G., 2006, “Bounded Rationality in Industrial Organization,” in *Advances in Economics and Econometrics: Theory and Applications*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press, Cambridge.

European Commission (EC), 2009, “On the Follow Up in Retail Financial Services to the Consumer Markets Scoreboard,” Commission Staff Working Document 1251, available at: http://ec.europa.eu/consumers/rights/docs/swd_retail_fin_services_en.pdf (accessed 19 October 2012).

European Commission (EC), 2011, “Proposal for a Directive on Markets in Financial Instruments repealing Directive 2004/39/EC (MiFID II),” available at: http://ec.europa.eu/internal_market/securities/isd/mifid_en.htm (accessed 11 July 2012).

European Commission (EC), 2012, “Proposal for a Directive on Insurance Mediation repealing Directive 2002/92/EC (IMD2),” available at: http://ec.europa.eu/internal_market/insurance/consumer/mediation/index_en.htm (accessed 15 July 2012).

European Insurance and Occupational Pensions Authority (EIOPA), 2011, “EIOPA report on the fifth Quantitative Impact Study (QIS5) for Solvency II,” available at: https://eiopa.europa.eu/fileadmin/tx_dam/files/publications/reports/QIS5_Report_Final.pdf (Accessed 11 May 2012).

Esty, B.C., 1997, “Organizational Form and Risk Taking in the Savings and Loan Industry,” *Journal of Financial Economics* 44 (1), 25–55.

Fama, E.F., and M.C. Jensen, 1983, “Agency Problems and Residual Claims,” *Journal of Law and Economics* 26 (2), 327-349.

Fecher, F., D. Kessler, S. Perelman, and P. Pestieau, 1993, “Productive Performance of the French Insurance Industry,” *Journal of Productivity Analysis* 4 (1-2), 73–93.

Fecher, F., S. Perelman, and P. Pestieau, 1991, “Scale Economies and Performance in the French Insurance Industry,” *Geneva Papers on Risk and Insurance* 16 (60), 315-326.

Fenn, P., D. Vencappa, S. Diacon, P. Klumpes, and Chr. O’Brien, 2008, “Market Structure and the Efficiency of European Insurance Companies: A Stochastic Frontier Analysis,” *Journal of Banking and Finance* 32 (1), 86–100.

Financial Services Authority (FSA), 2006a, “Advised and Non-Advised Sales - Insurance Factsheet”.

Financial Services Authority (FSA), 2006b, “A Guide to Mystery Shopping in Retail Financial Services”.

Financial Services Authority (FSA), 2009, "Distribution of Retail Investments: Delivering the RDR," Consultation Paper 09/18.

Financial Services Authority (FSA), 2012, "Guidance Consultation – Risks to Customers from Financial Incentives".

Financial Services Authority (FSA), 2013, "Payment Protection Insurance," available at: http://www.fsa.gov.uk/consumerinformation/product_news/insurance/payment_protection_insurance_ (accessed 8 February 2013).

Friedman, M., 1957, *A Theory of the Consumption Function*. Princeton University Press, Princeton.

Gatzert, N., and H. Schmeiser, 2008, "Combining Fair Pricing and Capital Requirements for Non-Life Insurance Companies," *Journal of Banking and Finance* 32 (12), 2589–2596.

Geneva Association, 2010, "Systemic Risk in Insurance - An Analysis of Insurance and Financial Stability".

Gerdtham, U.G., Johannesson, M., Lundberg, L., and D. Isacson, 1997, "A Note on Validating Wagstaff and van Doorslaer's Health Measure in the Analysis of Inequalities in Health," *Journal of Health Economics* 18 (1), 117-124.

Gorter, J.K., 2012, "Commission Bans and the Source and Quality of Financial Advice," Working Paper No. 350. De Nederlandsche Bank, Amsterdam.

Gorter, J.K., and J.A. Bikker, 2013, "Investment Risk Taking by Institutional Investors," *Applied Economics* 45 (33), 4629-4640.

Gorter, J.K., J.P.A.M. Jacobs, and J. de Haan, 2008, "Taylor Rules for the ECB using Expectations Data," *Scandinavian Journal of Economics* 110 (3), 473-488.

Gorter, J.K., and P. Schilp, 2012, "Risk Preferences over Small Stakes: Evidence from Deductible Choice," Working Paper No. 338. De Nederlandsche Bank, Amsterdam.

Grinblatt, M., S. Titman, and R. Wermers, 1995, "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior," *American Economic Review* 85 (5), 1088–1105.

Guiso, L., and M. Paiella, 2006, "The Role of Risk Aversion in Predicting Individual Behavior," in *Insurance: Theoretical Analysis and Policy Implications*, ed. by P-A. Chiappori, and C. Gollier. MIT Press, Boston.

Guiso, L., and M. Paiella, 2008, "Risk Aversion, Wealth, and Background Risk," *Journal of the European Economic Association* 6 (6), 1109–50.

Guo, D., and R.A. Winter, 1997, "The Capital Structure of Insurers: Theory and Evidence," mimeo.

De Haan, L., and J. Kakes, 2010, "Are Non-Risk Based Capital Requirements for Insurance Companies Binding?," *Journal of Banking and Finance* 34 (7), 1618–1627.

De Haan, L., and J. Kakes, 2011, "Momentum or Contrarian Investment Strategies: Evidence from Dutch Institutional Investors," *Journal of Banking and Finance* 35 (9), 2245-2251.

Hansmann, H., 1985, "The Organization of Insurance Companies: Mutual versus Stock," *Journal of Law, Economics, and Organization* 1 (1), 125-154.

Hanweck, G.A., and A.M.B. Hogan, 1996, "The Structure of the Property-Casualty Insurance Industry," *Journal of Economics and Business* 48 (2), 141-155.

Harrison, G.W., and E.E. Rutström, 2008, "Risk Aversion in the Laboratory," in *Research in Experimental Economics*, Vol. 12, ed. by J.C. Cox, and G.W. Harrison. Emerald Group Publishing Limited, Bingley.

Hartog, J., A. Ferrer-i-Carbonell, and N. Jonker, 2002, "Linking Measured Risk Aversion to Individual Characteristics," *Kyklos* 55 (1), 3-26.

Hirao, Y., and T. Inoue, 2004, "On the Cost Structure of the Japanese Property-Casualty Insurance industry," *Journal of Risk and Insurance* 71 (3), 501-530.

Holt, C.A., and S.K. Laury, 2002, "Risk Aversion and Incentive Effects," *American Economic Review* 92 (5), 1644-1655.

Hotelling, H., 1929, "Stability in Competition," *Economic Journal* 39 (153), 41-57.

Hubert, P.J., 1967, "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. M. Le Cam, and J. Neyman. University of California Press, Berkeley.

Hung, A.A., N. Clancy, J. Dominitz, E. Talley, C. Berrebi, and F. Suvankulov, 2008, "Investor and Industry Perspectives on Investment Advisers and Broker-Dealers," Technical Report, Rand Institute for Civil Justice.

Inderst, R., and M. Ottaviani, 2009, Misselling through Agents, *American Economic Review* 99 (3), 883-908.

Inderst, R., and M. Ottaviani, 2012a, "Financial Advice," *Journal of Economic Literature* 50 (2), 494-512.

Inderst, R., and M. Ottaviani, 2012b, "How (Not) to Pay for Advice: A Framework for Consumer Financial Protection," *Journal of Financial Economics* 105 (2), 393-411.

Jensen, M.C., 1986, "Agency Costs of Free Cash Flows, Corporate Finance and Takeovers," *American Economic Review* 76 (2), 323-339.

Jensen, M.C., and W.H. Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* 3 (4), 305-360.

Kahneman, D., and A. Tversky, 1979, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (2), 263-291.

Kahneman, D., and A. Tversky, 1992, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5 (4), 297-323.

Kakes, J., 2008, “Pensions in a Perfect Storm: Financial Behaviour of Dutch Pension Funds (2002-2005),” *Applied Financial Economics Letters* 4 (1), 29–33.

Van Kleef, R.C., K. Beck, W.P.M.M. van de Ven, and R.C.J.A. van Vliet, 2008, “Risk Equalization and Voluntary Deductibles: A Complex Interaction,” *Journal of Health Economics* 27 (2), 427-443.

Kőszegi, B., and M. Rabin, 2006, “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics* 121 (4), 1133-1165.

Kőszegi, B., and M. Rabin, 2007, “Reference-Dependent Risk Attitudes,” *American Economic Review* 97 (4), 1047-1073.

Kreijger, G., 2009, “Losses prompt Dutch Pension Funds to cut Equity Risk,” available at: http://www.reuters.com/article/2009/01/30/dutch-funds-abp-idUKLNE50T00320090130 (accessed 16 September 2012).

Lakonishok, J., A. Shleifer, and R.W. Vishny, 1992, “The Impact of Institutional Trading on Stock Prices,” *Journal of Financial Economics* 32 (1), 23–43.

Lamm-Tennant, J., and L.T. Starks, 1993, “Stock versus Mutual Ownership Structure,” *Journal of Business* 66 (1), 29-46.

Lee, S.J., D. Mayers, and C.W. Smith Jr., 1997, “Guarantee Funds and Risk Taking: Evidence from the Insurance Industry,” *Journal of Financial Economics* 44 (1), 3–24.

Leibenstein, H., 1966, “Allocative Efficiency versus “X-efficiency”,” *American Economic Review* 56 (3), 392–415.

Leibowitz, M., and A. Weinberger, 1982, “Contingent Immunization: Part 1: Risk Control Procedures,” *Financial Analysts Journal* 38, 17-31.

Lewis, T., and D. Sappington, 1994, “Supplying Information to Facilitate Price Discrimination,” *International Economic Review* 35 (2), 309-327.

Liebenberg, A.P., and D.W. Sommer, 2008, “Effects of Corporate Diversification: Evidence from the Property-Liability Insurance Industry,” *Journal of Risk and Insurance* 75 (4), 893-919.

Loewenstein, G., 1999, “Experimental Economics from the Vantage-Point of Behavioural Economics,” *Economic Journal* 109(453), F25-F34.

Lucas, D.J., and S.P. Zeldes, 2009, “How should Public Pension Plans invest?,” *American Economic Review: Papers & Proceedings* 99 (2), 527–532.

Mahlberg, B., and T. Url, 2000, “The Transition to the Single Market in the German Insurance Industry,” *Working Paper* 131. Austrian Institute of Economic Research, Vienna.

Malmendier, U., and D. Shantikumar, 2007, “Are Small Investors Naïve about Incentives?,” *Journal of Financial Economics* 85 (2), 457-489.

Mayers, D., and C.W. Smith Jr., 1981, “Contractual Provisions, Organizational Structure, and Conflict Control in Insurance Markets,” *Journal of Business* 54 (3), 407-434.

Mayers, D., and C.W. Smith Jr., 1987, "Corporate Insurance and the Underinvestment Problem," *Journal of Risk and Insurance* 54 (1), 45–54.

Mayers, D., and C.W. Smith Jr., 1988, "Ownership Structure across Lines of Property-Casualty Insurance," *Journal of Law and Economics* 31 (2), 351-378.

Mayers, D., and C.W. Smith Jr., 1992, "Executive Compensation in the Life Insurance Industry," *Journal of Business* 65 (1), 51–74.

Mayers, D., and C.W. Smith Jr., 1994, "Managerial Discretion, Regulation and Stock Insurer ownership structure," *Journal of Risk and Insurance* 61 (4), 638-655.

Meador, J.W., H.E. Ryan, Jr., and C.D. Schellhorn, 2000, "Product Focus versus Diversification: Estimates of X-efficiency for the U.S. life Insurance Industry," in *Performance of financial institutions: efficiency, innovation, regulation*, ed. by P.T. Harker, and S.A. Zenios. Cambridge University Press, New York.

Mester, L.J., 1989, "Testing for Expense Preference Behavior: Mutual versus Stock Savings and Loans," *Rand Journal of Economics* 20 (4), 483-498.

Meyer, M., P. Milgrom, and J. Roberts, 1992, "Organizational Prospects, Influence Costs, and Ownership Changes," *Journal of Economics & Management Strategy* 1 (1), 9-35.

Meyers, G.G., 1989, "Analysis of the Capital Structure of an Insurance Company," *Proceedings of the Casualty Actuarial Society* 76, 147-171.

Minister van Financiën, 2011, "Verdere Uitwerking Beleid Provisieverbod Complexe Producten," available at: http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken (accessed 16 May 2012).

Moscarini, G., and M. Ottaviani, 2001, "Price Competition for an Informed Buyer," *Journal of Economic Theory* 101 (2), 457-493.

Oosenbrug, A., 2007, "Insurance in the Netherlands: Market Structure and Recent Developments," in: *Handbook of international insurance: between global dynamics and local contingencies*, ed. by J.D. Cummins, and B. Venard. Springer, New York.

Pástor, L., and R.F. Stambaugh, 2012, "Are Stocks Really Less Volatile in the Long Run?," *Journal of Finance* 67 (2), 431-478.

Ponds, E.H.M., and B. Van Riel, 2009, "Sharing Risk: the Netherlands' New Approach to Pensions," *Journal of Pension Economics and Finance* 8 (1), 91-105.

Puelz, R., and A. Snow, 1994, "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market," *Journal of Political Economy* 102 (2), 236-257.

Rabin, M. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica* 68 (5), 1281-1292.

Rabin, M., and R.H. Thaler, 2001, "Anomalies: Risk Aversion," *Journal of Economic Perspectives* 15 (1), 219-232.

Rauh, J., 2009, "Risk shifting versus Risk Management: Investment Policy in Corporate Pension Plans," *Review of Financial Studies* 22 (7), 2687-2733.

Rees, R., and E. Kessner, 1998, "Regulation and Efficiency in European Insurance Markets," *Economic Policy* 14 (29), 365-392.

Van Rooij, M.C.J., Kool, C.J.M., and H.M. Prast, 2007, "Risk-Return Preferences in the Pension Domain: Are People Able to Choose?," *Journal of Public Economics* 91 (3-4), 701-722.

Saito, K., 2006, "Testing for Asymmetric Information in the Automobile Insurance Market Under Rate Regulation," *Journal of Risk and Insurance* 73 (2), 335-356.

Sharpe, W.F., 1976, "Corporate Pension Funding Policy," *Journal of Financial Economics* 3, 183-193.

Shleifer, A., Vishny, R., 1992, "Liquidation Values and Debt Capacity: A Market Equilibrium Approach," *Journal of Finance* 47 (4), 1343–1366.

Smith, C., and R. Stulz, 1985, "The Determinants of Firms' Hedging Policies," *Journal of Financial and Quantitative Analysis* 20 (4), 391-405.

Smith, B.D., and M.J. Stutzer, 1990, "Adverse Selection, Aggregate Uncertainty, and the Role for Mutual Insurance Contracts," *Journal of Business* 63 (4), 493-510.

Sommer, D.W., 1996, "The Impact of Firm Risk on Property-Liability Insurance Prices," *Journal of Risk and Insurance* 63 (3), 501–514.

Spierdijk, L., J.A. Bikker, and P. Van der Hoek, 2012, "Mean Reversion in International Stock Markets: An Empirical Analysis of the 20th Century," *Journal of International Money and Finance* 31 (2), 228-249.

Stiglitz, J.E., 2012, Rethinking Macroeconomics: What Failed and How to Repair It, *Journal of the European Economic Association* 9 (4), 591-645.

Sundaresan, S., and F. Zapatero, 1997, "Valuation, Optimal Asset Allocation and Retirement Incentives of Pension Plans," *Review of Financial Studies* 10 (3), 631–660.

Suret, J.-M., 1991, "Scale and Scope Economies in the Canadian Property and Casualty Insurance Industry," *The Geneva Papers on Risk and Insurance* 16 (59), 236-256.

Swiss Re, 1996, "Deregulation and Liberalization of Market Access: the European Insurance Industry on the Threshold of a new Era in Competition," *Sigma* 7.

Swiss Re, 1999, "Are Mutual Insurers an Endangered Species?," *Sigma* 4.

Swiss Re, 2000, "Europe in Focus: Non-Life Markets Undergoing Structural Change," *Sigma* 3.

Swiss Re, 2006, "World Insurance in 2005," *Sigma* 5.

Swiss Re, 2012, "World Insurance in 2011," *Sigma* 3.

Sydnor, J., 2010, "(Over)insuring Modest Risks," *American Economic Journal: Applied Economics* 2 (4), 177–199.

Teece, D.J., 1980, "Economies of Scope and the Scope of the Enterprise," *Journal of Economic Behavior and Organization* 1 (3), 223–247.

Teppa, F., and C. Vis, 2012, "The CentERpanel and the DNB Household Survey: Methodological Aspects," *DNB Occasional Study* 10 (4).

Toivanen, O., 1997, "Economies of Scale and Scope in the Finnish Non-Life Insurance Industry," *Journal of Banking and Finance* 21 (6), 759-779.

Treynor, J., 1977, "The Principles of Corporate Pension Finance," *Journal of Finance* 32, 627-638.

Turner, A., 2011, "Mansion House Speech," text available at: http://www.fsa.gov.uk/library/communication/speeches/2011/1020_at.shtml (accessed 25 January 2012).

United Nations Conference on Trade and Development (UNCTAD), 1964, *Proceedings of the United Nations Conference on Trade and Development*, First session, Vol. I, Final Act and Report, Annex A.IV.23.

Van de Ven, W.P.M.M., and F.T. Schut, 2008, "Universal Mandatory Health Insurance in The Netherlands: A Model for the United States?," *Health Affairs* 27 (3), 771-781.

Verbond van Verzekeraars, 2010, "Verzekeraars Wijzen op Mogelijkheid Aanpassing Beleggingspolis," available at: http://www.verzekeraars.nl (accessed 11 December 2012).

Verbond van Verzekeraars, 2012, "Verzekerd van Cijfers 2012," available at: http://www.verzekeraars.nl (accessed 19 October 2012).

Wagstaff, A., and E. van Doorslaer, 1994, "Measuring Inequalities in Health and in the Presence of Multiple-Category Morbidity Indicators," *Health Economics* 3 (4), 281-289.

White, H., 1980, "Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48 (4), 817-838.

Williamson, O.E., 1963, "Managerial Discretion and Business Behavior," *American Economic Review* 53 (5), 1032-1057.

Yuengert, A.W., 1993, "The Measurement of Efficiency in Life Insurance: Estimates of a Mixed Normal-Gamma Error Model," *Journal of Banking and Finance* 17 (2-3), 483-496**.**