MRIC | Working Papers

www.mric.lmu.de/research/wp

Credible or Biased?
An Analysis of Insurance Product Ratings in Germany

Patricia Born

Stephanie Meyr

Sharon Tennyson

Munich Risk and Insurance Center

Working Paper 31

April 29, 2015

Credible or Biased? An Analysis of Insurance Product Ratings in Germany
Patricia Born, Stephanie Meyr and Sharon Tennyson
MRIC Working Paper No. 31
April 2015

ABSTRACT

Product market transparency suffered from relaxed regulation following deregulation of the European insurance market in 1994. Instruments such as insurance product ratings can contribute to foster consumers' orientation in such market environments. However, the capacity of these ratings to promote market transparency and consumer awareness depends critically on whether they are credible. This article provides the first empirical investigation of insurance product ratings, with an emphasis on the potential sources of bias that could undermine rating credibility. The analysis employs a panel data set containing ratings for German disability insurance products from two rating agencies over a fifteen year period. Using the existing literature on other rating types as a guide, we test a series of hypotheses regarding factors that may explain the variation in rating outcomes over time and across rating agencies. Even if conspicuous differences between rating agencies can be revealed, our results suggest no major concerns regarding the credibility of insurance product ratings.

Patricia Born
Center for Insurance Research
Department of Risk Management/Insurance, Real Estate and Legal Studies
Florida State University - College of Business
821 Academic Way
Tallahassee, FL 32306-1110
United States of America
pborn@cob.fsu.edu

Stephanie Meyr
Munich Risk and Insurance Center (MRIC)
Munich School of Management
Ludwig-Maximilians-Universität in Munich
Schackstraße 4/III
80539 Munich
Germany
meyr@bwl.lmu.de

Sharon Tennyson
Cornell Institute for Public Affairs (CIPA)
Department of Policy Analysis and Management (PAM)
Cornell University
292 Caldwell Hall
Ithaca, NY 14853
United States of America
sharon.tennyson@cornell.edu

Introduction

Insurance products, especially those for life, health and disability coverages, are widely recognized for their complexity. Transparency of product features is important to ensuring optimal market outcomes by enabling consumers to accurately assess their need for coverage and their willingness to pay for certain features. A variety of regulatory measures are used in insurance markets around the globe to address this issue, including in some cases explicit regulation of product features. However, markets regulated to such an extent lose the potential benefits of free competition. Information markets, for example the provision of product ratings, are an alternative approach to promoting transparency in unregulated product markets. This has been the approach used in Europe since European Union Directives deregulated insurance product markets in 1994. Insurance product features are no longer subject to regulatory prior approval before market launch, but consumers are able to compare the quality of insurance products using product ratings provided by government and private raters. Ideally, this combination of free competition accompanied by transparency-enhancing market mechanisms could lead to more optimal market outcomes.

From a theoretical standpoint Akerlof (1970) first demonstrated how information asymmetries about product quality could bring severe efficiency problems to a market, and could lead to market collapse in the worst-case. Shortly after this seminal article, quality certification (e.g. ratings) by information intermediaries was first discussed as a remedy for these asymmetric information problems (Viscusi, 1978; Leland, 1979). The net welfare effects of adding quality certification to a market depend critically, however, on the quality of the certifications themselves. In a recent review of the literature on this subject, Dranove and Jin (2010) argue that two failures of certifications may reduce their usefulness in improving market performance: bias and imprecision. The theoretical literature shows that rating imprecision is mainly a problem in markets with a single rating agency, and that imprecision stems from the rating monopolist's desire to capture rents (Lizzeri, 1999; Doherty et al., 2012). The literature also shows, however, that rating bias can be affected by a rating market's governance, structure, contractual relationships, information flows, or uses of ratings.[1] Thus, biased ratings create more potential for deadweight loss in rating markets than do imprecise ratings, and we focus on rating bias in this study.[2]

Based on the unique case of the development of a market for insurance product ratings after insurance deregulation in Germany, this study aims to evaluate the product rating market with a special focus on sources of potential rating bias. One important source of bias in privately-provided ratings stems from the raters' financial incentives: the opportunity to make money from providing a rating now as well as the opportunity for future revenue from providing subsequent ratings. In many market settings,

[1] See the discussion in Dranove and Jin (2010), and papers cited in Meyr and Tennyson (2015) for more details.

[2] Another reason we focus on rating bias is that – unlike for financial ratings – there are no obvious sources of data to permit a quantitative assessment of the accuracy of product ratings.

financial incentives are structured to reward upwardly biased ratings. We propose in this study that insurance products are likely to be over-rated if business relationships with the issuing insurers are expected to be particularly beneficial for the rating agency. We evaluate, empirically, this relationship and several other potential sources of bias.

This study contributes to the literature in two main ways. Firstly, it provides new evidence on the validity of product ratings, albeit in a specific context. Secondly, the evidence produced here may inform the academic debate on insurance market transparency and product regulation more generally. There is currently no similar market for insurance product ratings in the U.S., for example, although the products are no less complex than in Germany.[3] If an information market can provide valid, unbiased ratings of products, even if the agencies in this market operate on a for-profit basis, the need for government intervention in the form of strict product regulation may be reduced (see Meyr and Tennyson, 2015).

The remainder of the paper proceeds as follows: The next section provides background on the market for insurance product ratings in Germany. We explain the factors that are considered when rating a product, and describe the rating market structure. In section 3 we describe the dataset on product ratings and the supplemental data used to complete our analysis. Section 4 develops hypotheses to be tested, drawing on previous literature and the institutional features of the rating market. Section 5 discusses empirical methodology and presents results. A final section discusses our findings and provides policy implications.

Background

The Product Rating Market

Insurance product ratings are third party quality certificates for insurance products. That is, product ratings are external assessments of the quality of a specific insurance contract that an insurer provides, based on features such as the terms and conditions of coverage, clarity of sales documents and the application form and process. Such ratings are distinct from financial strength ratings or credit ratings, which focus on the financial and other quality aspects at the enterprise level using balance sheet, income, and operating performance data. There are some common considerations in the two types of ratings, of course. For example, product ratings additionally factor in enterprise characteristics that bear on the insurer's fitness as the provider of a specific insurance product; and, financial strength ratings take account of product related factors such as claims payment processes and policies, but not with reference to specific products.[4]

[3] In the U.S., the Health Plan Report cards provided by the National Committee for Quality Assurance might come closest to the idea of insurance product ratings as provided in the German market.

[4] See for example the "Guide to Best's Financial Strength Ratings".

Product rating agencies began to enter the German insurance market in 1995, the year after implementation of the European Union's Third Insurance Directives which deregulated insurance markets,[5] and have become an important fixture in the market. At least two studies show that insurance product ratings are highly recognized within the German market. According to Romeike (2004)[6] consumers are very likely to consult ratings before choosing an insurance company (72%). A more recent survey by Assekurata (2006) suggests that more than 80% of consumers at least occasionally consult product ratings when they search for information on insurance products. Further, the vast majority of consumers (74%), brokers (78%) and even insurance companies (79%) in that survey responded that they trust the ratings. Rating seals which identify the rating(s) of a product are typically used as one component of insurers' advertising – they are shown prominently in brochures and on websites – and brokers use them to identify products they prefer to sell as well as to justify their advice. Additionally, product ratings are published in consumer magazines and, more recently, these ratings are commonly used in online product comparisons.

Insurance product ratings in Germany are provided by several private agencies as well as by one government foundation.[7] Private and public agencies show significant differences in objectives, target groups for ratings, and revenue sources. While private agencies' primary goal is profit maximization, more consumer-orientated objectives direct the actions of the government rating agency.[8] In contrast to credit ratings, however, the insurance product ratings are not commissioned or paid for by the insurance companies. Rating agencies assess the products' quality on their own initiative and choice.[9] Nonetheless, some product ratings rely in part on internal information provided by the insurer (so-called interactive ratings), and thus do require the insurer's cooperation to produce. Many product ratings use only publicly-available information (so-called PI-ratings), however.

In this paper, we concentrate on ratings for occupational disability products. These products provide coverage for loss of earnings caused by invalidity. Although this is only a subset of product ratings, it provides a useful case study. First, occupational disability plans account for significant proportions of the life insurance business in Germany, as disability risk was privatized by social security reforms in 2001. Ratings on these products are quite common and provided by almost every rating agency on the German market, due to product complexity and the importance of these products for consumers. Additionally, ratings for these products have been provided continuously over the years, while

5 See Berry-Stölzle and Born (2012) for a description of the deregulation in Germany.

6 This survey is published by Swiss Re in 1997. We do not have access to the original study.

7 Rating agencies providing insurance product ratings are not affected by the European regulation of rating agencies.

8 The purpose of the governmental foundation Stiftung Warentest outlined in their statutes reads as follows: "The foundation works selflessly; it does not primarily pursue its own financial interests. Purpose of the foundation is to foster consumer protection…".

9 We are aware of just one exception which is the rating by Versicherungsforen Leipzig in cooperation with TÜV Saarland. This rating, which may be purchased by insurers, is the newest on the market with their first rating issued in 2013.

fundamental product characteristics have remained quite stable. This makes examination of ratings over time a meaningful exercise. Moreover, the characteristics of ratings for these products should be representative of those for other insurance products since the rating systems consider factors that are also used in rating life, health and property-casualty insurance products. This is to be expected, because occupational disability insurance combines characteristics of life insurance products with the more complex contractual terms regarding obligations and conditions for claim payments that are seen in health and property-casualty insurance.

Our analysis is based on the ratings of two important agencies: Morgen & Morgen GmbH, a private rating agency and Stiftung Warentest, a government foundation that provides the so-called Finanztest ratings. The Morgen & Morgen ratings are interactive ratings, since one subset of rating factors is obtained from a survey of insurers. The Finanztest ratings are PI-ratings, but Stiftung Warentest relies on insurance companies to provide the data, and thus the rating is effectively interactive.

The current rating system for occupational disability products of Morgen & Morgen has been in place since 2004.[10] Stiftung Warentest has not maintained a consistent approach over time with their Finanztest ratings. In some years the Finanztest rating focuses on particular aspects or target groups of occupational disability insurance products; in other years, the foundation decided to rate products which could provide alternative solutions to cover invalidity risks.[11] For this rating agency we restrict our analyses to years in which they rate occupational disability products and target the majority of consumers, and drop years with special focuses (2002, 2012).

Product Rating Data

We construct a hand-collected dataset of 4,244 observations for Morgen & Morgen ratings in the years 1999 to 2013 and 1,004 observations for Finanztest ratings from years between 2000 and 2013. The number of products rated in each year, for each agency, is shown in Table 1.

[10] Detailed information on the Morgen & Morgen rating system is available on their website (www.morgenundmorgen.de).

[11] Detailed information on the applied rating system is outlined in the Finanztest magazines the ratings are published. In 2011 and 2013, Stiftung Warentest considered only each insurer's least expensive contracts. For 2000-2001, 2009-2010, and 2003-2007 the agency considered occupational disability contracts in combination with term life insurance and in 2008 examined offers for stand-alone contracts instead. In 2004 the Finanztest rating contained both combined products and stand-alone tariffs. These changes are interesting but do not influence our investigation as our models compare ratings at the product level.

Table 1: Sample Characteristics: Rating Data

Number of observations	Year of Rating														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Morgen & Morgen	246	183	188	178	193	229	265	303	342	348	361	352	344	356	356
Finanztest	0	110	106	0	93	138	89	83	85	55	78	39	54	0	74
Total	246	293	294	178	286	367	354	386	427	403	439	391	398	356	430

The ratings are issued for 873 occupational disability products in total. Figure 1 displays the mean, median, minimum and maximum numbers of products rated by at least one agency on the insurer-year-level. As the rating agencies aim to provide a comprehensive reflection of the considered market, this provides evidence for an increasing number of rated products per insurer over time. This suggests an increase in complexity of the market over time.

Figure 1: Number of Rated Products per Insurer

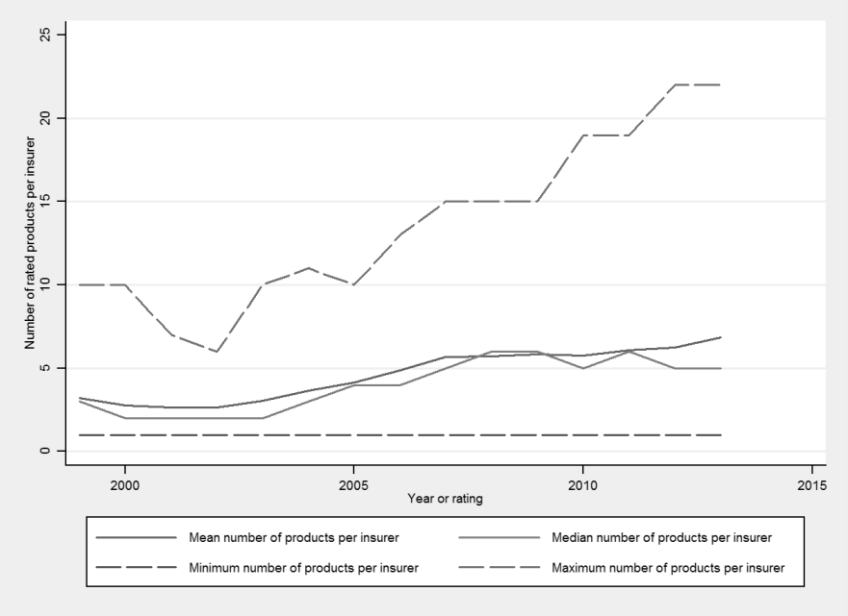
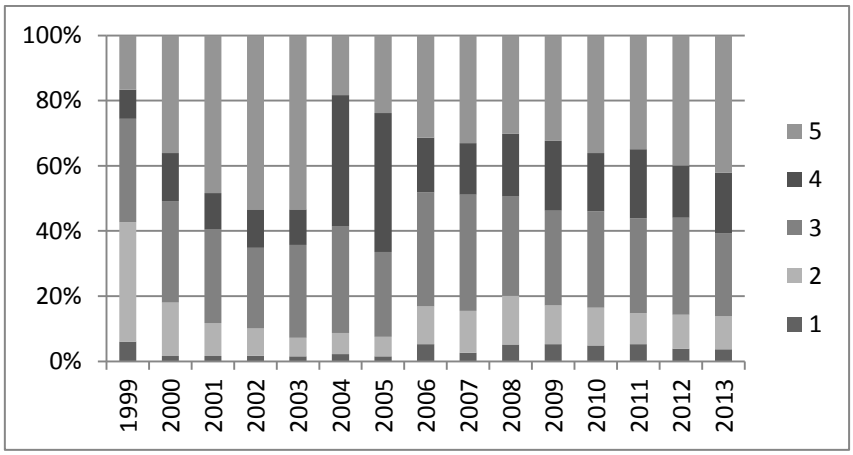


Figure 2 shows the distribution of ratings published by Morgen & Morgen for our sample period. Morgen & Morgen ratings are issued on a five-point-scale increasing with product quality. While in the beginning of our sample it appears that they awarded an increasing number of highest ratings (5)

over time, Morgen & Morgen changed its rating system in 2003 with the consequence that the proportion of these highest ratings is less than 40 percent after that.[12]

Figure 2: Distribution of Morgen & Morgen Ratings, 1999-2013

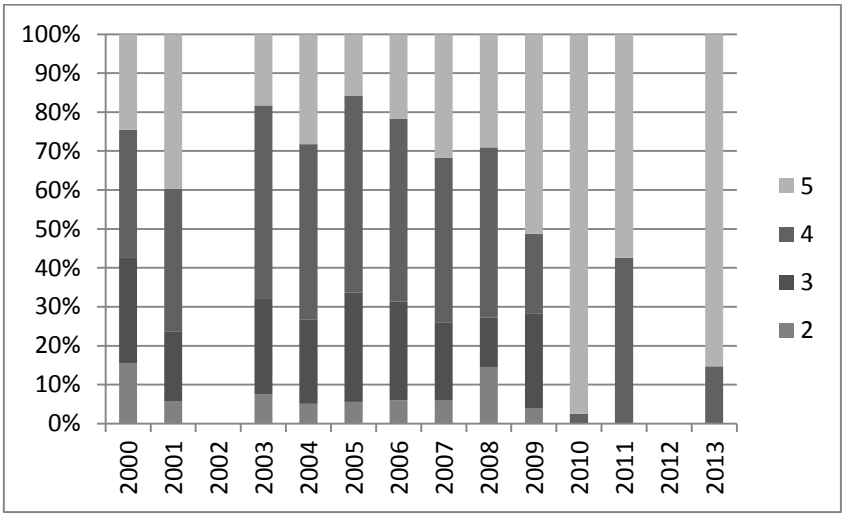


Similarly, we show the distribution of product ratings by Finanztest for the period 2000 to 2013 in Figure 3. Finanztest ratings are originally published on a continuous scale from 0.5 for the best product quality to 5.5 for the worst product quality. However, to facilitate readability Stiftung Warentest clusters their numeric ratings into five quality groups ("very good", "good", "satisfactory", "sufficient" and "defective"). We translate these five quality groups into a five-point-scale as used by Morgen & Morgen.[13] The Finanztest data suggest an overwhelming proportion of highest ratings since 2009. In 2010, for example, only one product received a rating of "4", while all others (38) received a "5". This trend in Finanztest ratings distribution towards good ratings does not necessary mean biased ratings since it could be caused by changes in the portfolio of products rated each year.

[12] The current rating system for occupational disability products of Morgen & Morgen has been in place since 2004. Before 2004, their rating system was mainly focused on the comparison of terms and conditions.

[13] This method of ratings scale transformation follows the approach used by Pottier and Sommer (1999). This one-to-one mapping is appropriate because the descriptions of the Morgen & Morgen rating categories almost exactly correspond to the Finanztest category descriptions (e.g. Morgen & Morgen's category three is "average" where Stiftung Warentest refers to this as "satisfactory").

Figure 3: Distribution of Finanztest Ratings, 2000-2013



The contemporaneous (annual) Spearman correlation between Morgen & Morgen and Finanztest ratings is 0.677 and ranges from 0.491 to 0.834 across years. Table 2 provides summary data on the relationship between ratings provided by Morgen & Morgen and Stiftung Warentest, for products rated by both agencies. Morgen & Morgen publishes their ratings usually in April whereas Finanztest ratings are normally published in July. The first panel compares the Finanztest ratings to the Morgen & Morgen ratings published three months earlier. The second panel compares the Morgen & Morgen ratings to the Finanztest ratings published in the year before. For the sample period as a whole, the majority of ratings provided by each agency are the same as the earlier published ratings by the other agency: 56 percent of ratings are equal when comparing Morgen & Morgen ratings with following Finanztest ratings, and 57 percent of ratings are equal when comparing Finanztest ratings with Morgen & Morgen ratings following in the next period. There is also no strong pattern in the direction of differences: 24 percent of Finanztest ratings are lower than the previous Morgen & Morgen rating and 25 percent are higher; 21 percent of Morgen & Morgen ratings are lower than the previous Finanztest rating and 23 percent are higher.

Table 2: Differences between Morgen & Morgen and Finanztest Ratings

Comparison of earlier MM ratings and following FT ratings of the same year

	Year of rating												
	2000	2001	2003	2004	2005	2006	2007	2008	2009	2010	2011	2013	Total
MM_t smaller than FT_t by 2	3 (3.5)	1 (1.1)	0 (0.0)	0 (0.0)	1 (1.4)	0 (0.0)	1 (1.3)	0 (0.0)	2 (2.7)	1 (2.9)	1 (2.0)	3 (4.5)	13 (1.5)
MM_t smaller than FT_t by 1	14 (16.1)	14 (15.1)	5 (6.4)	33 (28.7)	12 (16.4)	15 (21.4)	17 (22.1)	12 (25.5)	22 (30.1)	7 (20.6)	8 (16.3)	17 (25.4)	176 (20.4)
MM_t and FT_t equal	48 (55.2)	59 (63.4)	31 (39.7)	60 (52.2)	36 (49.3)	37 (52.9)	46 (59.7)	21 (44.7)	42 (57.5)	26 (76.5)	29 (59.2)	45 (67.2)	480 (55.6)
MM_t exceeding FT_t by 1	22 (25.3)	19 (20.4)	42 (53.9)	20 (17.4)	23 (31.5)	17 (24.3)	12 (15.6)	14 (29.8)	6 (8.2)	0 (0.0)	11 (22.5)	2 (3.0)	188 (21.8)
MM_t exceeding FT_t by 2	0 (0.0)	0 (0.0)	0 (0.00)	2 (1.7)	1 (1.4)	1 (1.4)	1 (1.3)	0 (0.0)	1 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.7)
Total number of commonly rated products	87	93	78	115	73	70	77	47	73	34	49	67	863

Comparison of earlier FT ratings and following MM ratings of the next year

	Year of FT rating												
	2000	2001	2003	2004	2005	2006	2007	2008	2009	2010	2011	2013	Total
MM_t smaller than FT_{t-1} by 4	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (4.6)	1 (0.1)
MM_t smaller than FT_{t-1} by 3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	7 (31.8)	7 (0.9)
MM_t smaller than FT_{t-1} by 2	1 (1.1)	1 (1.2)	1 (1.3)	2 (1.7)	2 (2.9)	0 (0.0)	2 (2.9)	0 (0.0)	2 (3.0)	1 (2.9)	2 (4.2)	8 (36.4)	22 (2.8)
MM_t smaller than FT_{t-1} by 1	9 (9.6)	9 (10.5)	10 (13.2)	23 (20.0)	13 (18.8)	13 (18.8)	16 (22.9)	10 (21.7)	16 (23.9)	9 (26.5)	5 (10.4)	1 (4.6)	134 (16.9)
FT_{t-1} and MM_t equal	48 (51.1)	53 (61.6)	51 (67.1)	66 (57.4)	35 (50.7)	37 (55.1)	37 (54.3)	23 (50.0)	40 (59.7)	24 (70.6)	29 (60.4)	5 (22.7)	448 (56.4)
MM_t exceeding FT_{t-1} by 1	28 (29.8)	22 (25.6)	14 (18.4)	22 (19.1)	17 (24.6)	17 (24.6)	14 (20.0)	13 (28.3)	8 (11.9)	0 (0.0)	12 (25.0)	0 (0.0)	167 (21.0)
MM_t exceeding FT_{t-1} by 2	8 (8.5)	1 (1.2)	0 (0.0)	2 (1.7)	2 (2.9)	1 (1.5)	0 (0.0)	0 (0.0)	1 (1.5)	0 (0.0)	0 (0.0)	0 (0.0)	15 (1.9)
Total number of commonly rated products	94	86	76	115	69	68	69	46	67	34	48	22	794

Numbers in brackets indicate percentage shares.

However, the data indicate a change in the patterns over time. Comparing the first six sample years to the second six, there is an increasing tendency for Finanztest ratings to match the previous Morgen & Morgen rating in the latter period. For years 1999-2005, 46 percent of Finanztest ratings are the same as the rating provided by Morgen & Morgen in the previous publication, but for years 2006-2012 Finanztest ratings match the previous Morgen & Morgen rating 58 percent of the time. The data also show that this is due to a reduction in Stiftung Warentest's propensity to provide a lower rating than Morgen & Morgen: this occurred for 31 percent of products during 1999-2005 but for only 15 percent of products during 2006-2012. Thus, Finanztest ratings appear increasingly to match those of Morgen & Morgen over time.

Development of Hypotheses

The literature on rating markets notes that the design of contractual relationships between rating agencies and rated companies and the consequential cash-flows are a major source of potential conflicts-of-interest for raters. One concern is collusion between rating agencies and the companies that they rate, especially in markets where ratings are solicited and paid for by the rated company. Since a rating agency's interest is to maximize profit, in their attempt to attract the maximum number

of products to rate they may have an incentive to offer upwardly biased initial ratings. If there are weak reputational penalties for inaccurate ratings, no incentives exist to correct the over-rating in subsequent periods (Strier, 2008).

Some empirical studies of credit ratings yield evidence of upward bias due to collusion. For example, Poon (2003) examines the effects of rating commissioning on credit ratings using 2-year panel data on ratings of 15 countries. Her results indicate that ratings are lower for unsolicited quality assessments, which suggests an upward bias in commissioned ratings due to collusion between rating agencies and rated companies.[14] Covitz and Harrison (2003) argue, however, that competition in rating markets will reduce the potential for bias due to collusion. Rating agencies are naturally led by an objective to gain and keep a high level of reputation. The ability of the users of ratings to evaluate rating agencies' credibility by comparing their ratings' quality grows with the number of competitors providing ratings. As a consequence, the importance of reputation increases with the level of competition in the rating market.

Market discipline from competition is not likely to be strong in the market for insurance product ratings because rating accuracy is difficult to measure, even with the passage of time. Unlike for bond or credit ratings, where ex-post performance measures of the rated instrument or firm are available (e.g., failure rate or market performance), information about an insurance product's "true" quality is nearly impossible to discern. Even though a greater number of ratings per product permit consumers of ratings to compare the recommendations of different raters, this provides only relative information about raters and product ratings. Thus, there is no guarantee that increased competition in the product rating market will improve rating accuracy.

Nonetheless, the potential for collusion between rating agencies and insurers is lower in the German market for insurance product ratings than in some other rating markets, because the product ratings are typically not commissioned by insurers and the rating agencies choose the products to rate. Both of these aspects are found to be potential remedies for rating bias by Bolton, Freixas and Shapiro (2012).

There remains a potential profit incentive for upward bias in product ratings if insurers are more likely to purchase the rating seals for more highly rated products.[15] Moreover, the lack of objective measures of rating accuracy means that the product rating market may also have weak reputational penalties for inaccurate ratings. These features could lead to upward bias in ratings and to little incentive for rating agencies to correct the upward bias over time. But although insurance companies do pay license fees for rating seals, the resulting revenues are quite small compared to other sources of revenue for the rating agencies (Meyr and Tennyson 2015). Beaver, Shakespeare and Soliman (2006) argue that

[14] Lower ratings for unsolicited quality assessments could also result from sample selection bias or the rating procedure applied. Nevertheless, there is some evidence of upward bias in commissioned ratings as regarding the Japanese market. Poon's (2003) results still hold when controlling for financial factors characterizing the rated companies.

[15] We do not observe price competition as the license fees for rating seals remained quite stable over time per agency.

collusion is an unlikely explanation for differences in bond ratings across raters when the proportion of revenues per customer is sufficiently small. Covitz and Harrison (2003) also mention diversification of revenue base as a potential remedy to conflicts-of-interest leading to upward ratings bias.

These arguments do not rule out the possibility that the insurance product ratings will be biased upward, but suggest that the features which provide strong incentives for systematic upward bias in ratings are not strongly present in this market. For this reason, we do not expect a strong upward pressure on product ratings overall. This does not equate to a lack of bias in the ratings, however. Product ratings are not required by law, as might be the case for some credit ratings or bond ratings. Thus, insurance product rating agencies depend greatly on customers' and brokers' awareness. Larger insurance companies are usually more familiar to customers and brokers, and can therefore make a greater contribution to increasing the rating agencies' prominence by the use of the rating seals in marketing, compared with smaller insurance companies. The desire to attract these larger insurers and keep them as customers could also provide incentives for inflated ratings. Based on this reasoning, we propose several related hypotheses regarding bias in insurance product ratings. We propose Hypothesis 1, expressed as the alternative to the null hypothesis of no effects, as follows:

H1: Larger insurance companies receive higher ratings per product, all other factors held constant.

However, because the literature on industrial organization and strategic management suggest that larger firm size may reflect higher productivity or other competitive advantages, it is possible that larger firms offer better products (see e.g., Barney 2001). Thus, the firm size effect alone cannot be interpreted as strong evidence of bias in insurance product ratings. We additionally expect that insurers which offer a larger product portfolio might receive higher ratings as they will potentially buy a greater number of rating seals.[16] Thus, we propose Hypothesis 2, expressed as the alternative to the null:

H2: Ratings per product increase with the number of products provided by an insurance company.

Finally, we consider the long-term effects of rating agency-insurer relationships. The longer the duration of the business relationships between rating agencies and insurers, the larger the potential incentives for upward bias as the agency does not want to endanger loyal sources of revenue. Rating improvements over time could equally result from insurers' learning about the requirements to receive high ratings. Our empirical models will account for that by incorporating sporadic changes in rating

[16] Insurer size and numbers of products provided by the insurer is not necessarily positively correlated. Smaller insurers might for example be specialists for a particular product type and therefore offer a broader spectrum. With Pearson's correlation coefficient being 0.3592 we do also not find a strong interrelationship between net premiums and number of products provided by an insurer in our data set.

systems as control variables. Specifically we suggest Hypothesis 3, expressed as the alternative to the null:

H3: Ratings per product increase with the number of periods a product has been rated before the current rating by one agency.

Rating bias could also arise from competitive pressure. Insurers' decisions on rating seal purchase and data provision could cause competition among rating agencies. Even if the rated companies' market power is not as high as for credit ratings, the intent to maximize the number of rating seals sold might still motivate rating agencies to adjust their ratings in dependence on competitors' assessments. As a consequence, ratings for a specific product might directly depend on prior ratings of other agencies. We therefore test the following Hypotheses 4a and 4b, expressed as the alternatives to the null:

H4a: Downgrades for a particular product from agency A are followed by non-downgrades by agency B in the next rating period.

H4b: Upgrades for a particular product from agency A are followed by upgrades by agency B in the next rating period.

Empirical Analysis of H1-H3

Data

For the empirical analysis we use the panel data set of product rating data, combined with insurance company data for the German life insurance market. Each observed product rating in our data set is matched with company specific information on the insurer that provides the rated product. Consequently, company-specific data is provided on 141 German life insurance companies for all years of our rating sample period. All insurer data were obtained from Bisnode, a private provider of data on financial company characteristics and financial performance measures. Appendix Table A1 presents summary statistics for these data.

We are unable to obtain financial data for all insurers that offer disability insurance products. As a result of merging the rating data with the company characteristics data, rating observations reduce to 3,383 by Morgen & Morgen and 802 by Stiftung Warentest if we restrict our considerations to ratings of products for which we have financial data for the insurer. For all empirical analysis that uses insurance company data in addition to rating data, we use the restricted number of observations.

Variables

We employ data on insurance company age, organizational form, ownership form, balance sheet and underwriting data; data on product offerings including average insured amount per contract; and

selected performance indicators including the loss ratio and the contract cancellation rate. Some variables are used to provide key information about the effects of provider-specific characteristics on product ratings. Other variables are used to control for differences in financial and organizational characteristics of the insurer, which may affect product ratings.

To test whether the likelihood of receiving more favorable ratings increases with company size as outlined in Hypothesis 1 our empirical models include the variable *log net premium.*[17] The suspicion of bias resulting from inflated ratings for big players on the insurance market could not be eliminated if we find evidence for a positive significant impact of *log net premium* on ratings. In that case Hypothesis 1 could be confirmed.

Hypothesis 2 follows a similar path of argumentation as it states a growing potential for bias for companies providing a greater product portfolio. To test Hypothesis 2 we incorporate the variable *number of rated products* into our models. This variable contains the number of the insurer's products rated by the considered rating agency in each year. To confirm Hypothesis 2 we expect to find a positive significant influence of this variable on a product rating.

Whether incentives to keep long-lasting business relationships lead to inflated ratings should be tested with Hypothesis 3. We incorporate the variable *number of years rated in a row* in our model. The variable counts the number of consecutive ratings for a particular product before the current rating. In order to confirm Hypothesis 3 we would again expect a positive and significant relationship between years rated and the rating.

In addition to these three key variables our model is complemented by a variety of firm characteristics as controls. We expect that older companies mostly provide well-known brands and might be advanced in product design and distribution. The variable *company age* is included for influences arising from companies' experience and establishment. Effects in both directions could be expected: either these companies see less need to bring their products into shape for receiving high ratings as they feel quite independent from rating effects, or the opposite is the case when rating agencies expect benefits of assigning high ratings to products of well established companies. The latter argument goes in line with our hypotheses on firm size and product portfolio.

As a second group of controls we take into consideration the legal form of an insurance company which could be stock, mutual and public-service enterprises, or an establishment of a foreign insurance company.[18] Indicator variables are incorporated into the model for *public*, *mutual* and *foreign* companies, and stock companies serve as the reference group. Special characteristics of the different legal forms regarding target groups, financing or decision-making might influence product design. As

[17] Results did not change when alternative measures for company size (total assets and equity capital) are applied.

[18] According to the national Insurance Supervision Act insurance companies in Germany are restricted to these four entrepreneurial forms. Besides stock and mutual insurers, companies established under public law have a long tradition as insurance providers in Germany.

just one example, mutual insurers are owned by policyholders and could be expected to design products of better quality compared to stock companies.

The *change in loss ratio* includes the relationship between incurred losses and earned premiums, and should serve as a control to incorporate the quality of underwriting on the one hand and the adequacy of pricing on the other hand. We expect positive effects on ratings. Lastly, our model also contains controls at the insurance company level to capture customer satisfaction and information on target group characteristics. First, with the variable *cancellation quota* we aim to proxy general consumer satisfaction with the insurer's products, e.g., due to service quality, product quality, or both. We expect that ratings are negatively associated with cancellations. Second, we incorporate the *average sum insured per contract* of an insurer. In Germany, insurance companies with big contracts in their portfolio are more likely to distribute their products via insurance brokers. Brokers should have a higher interest in ratings compared to exclusive agents, for example. Hence, the average size of contracts might have a positive influence on ratings as insurers care more about displaying rating seals and consequently designing their products with a special focus on rating fit.

Methodology

We test Hypotheses 1 to 3 using the sample of 4,185 unique product-year entities contained in the panel data set described above. The empirical analysis utilizes an ordered multinomial model.[19] Rating categories per product at year R_{it} will serve as dependent variable. Possible outcomes correspond to the rating scale and therefore can take ordinal values from 1 (poor quality) to 5 (very good quality). The model can be derived from the following latent variable model where $\mu_1$to μ_5 represent unknown thresholds for the observed rating categories:

$$R_{it} = \begin{cases} 1 \;\; if \;\; R_{it}^* < \mu_1 \\ 2 \;\; if \;\; \mu_1 \leq R_{it}^* < \mu_2 \\ 3 \;\; if \;\; \mu_2 \leq R_{it}^* < \mu_3 \\ 4 \;\; if \;\; \mu_3 \leq R_{it}^* < \mu_4 \\ 5 \;\; if \;\; R_{it}^* \geq \mu_4 \end{cases}$$

The estimation can be done using an ordered probit regression. The influence of insurance company size, number of products in its portfolio and length of business relationship before the current rating will be examined. We use clustered standard errors on the firm-level as bias due to within-firm correlation over time should be avoided. The ordered probit model can be described by the following equation:

[19] The model design is based on Blume et al. (1998).

$$\Pr(R_{it} = j|\theta) = \begin{cases} \Pr(\beta_0 + \beta' \mathrm{X}_{it} + \epsilon_{it} \geq \mu_4|\theta) & if\ j = 5 \\ \Pr(\mu_j > \beta_0 + \beta' \mathrm{X}_{it} + \epsilon_{it} \geq \mu_{j-1}|\theta) & if\ j = 4{,}3{,}2 \\ \Pr(\mu_1 > \beta_0 + \beta' \mathrm{X}_{it} + \epsilon_{it}|\theta) & if\ j = 1. \end{cases}$$

To control for time trends over the 15 years in our data set we additionally provide a model variation including year dummies.[20]

Estimation Results

The following set of estimates examines the effect of company size, number of rated products, and length of business relationship on ratings to test Hypotheses 1 to 3. Table 3 reports the results of the ordered probit estimation for the two rating agencies Morgen & Morgen and Stiftung Warentest.

Estimation results show that the effect of company size on the product rating is significant and positive for both rating agencies. This is consistent with Hypothesis 1. Counter to Hypothesis 2, the number of rated products for an insurance company is negatively associated with the rating, and this relationship is statistically significant for Morgen & Morgen ratings. The effect of the number of years a product has been rated in a row is contrary to Hypothesis 3 for Finanztest ratings, as it is negative and significant. Concerning Morgen & Morgen ratings, the effect is more in line with our hypothesis but is not statistically significant.

Thus, the overall estimation results are mixed with respect to Hypothesis 1-3 – that ratings will be upwardly biased for insurance companies from which the rating agencies expect to generate more revenue directly (through sales of certificates) or indirectly (through publicity from the rating seals). The only result that is clearly supportive of the hypothesized effects is a positive and significant relationship between net premiums of the insurer and the product rating. Thus, the evidence for upward bias in ratings is mixed.

[20] The year dummies also help control for potential effects on ratings of the rating system change by Morgen & Morgen in 2004. In an alternative specification, a single dummy variable representing the pre reform period was found to be a significant determinant of ratings. However, this change has no effect on the other variables of interest.

Table 3: Ordered Probit for Hypotheses 1 to 3

	Morgen & Morgen				Finanztest			
	Model IA:		Model IB:		Model IA:		Model IB:	
	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value
Full sample	**N = 3,225**		**N = 3,225**		**N = 766**		**N = 766**	
Log net premium	0.237	0.007***	0.238	0.020**	0.328	0.000***	0.258	0.003***
Number of rated products	-0.068	0.025**	-0.050	0.094*	-0.044	0.297	-0.114	0.024**
Number of years rated in a row	0.001	0.968	-0.042	0.289	-0.098	0.020**	-0.024	0.687
Age of company	0.006	0.093*	0.008	0.066*	0.006	0.159	0.008	0.039**
Mutual company	0.145	0.751	0.095	0.853	-0.813	0.053*	-0.930	0.017**
Public organization	-1.135	0.000***	-0.846	0.018**	0.944	0.001***	0.677	0.027**
Establishment of foreign insurer	0.742	0.024**	0.762	0.037**	1.534	0.000***	0.598	0.139
Change in loss ratio	0.038	0.468	-0.023	0.749	0.374	0.022**	0.261	0.021**
Cancellation ratio	0.019	0.508	0.034	0.264	0.003	0.958	0.061	0.330
Average sum insured per contract	0.019	0.009**	0.019	0.036**	0.035	0.000***	0.024	0.004***
Year dummies	No		Yes		No		Yes	

*** Indicates the difference from zero is statistically significant at the 1% confidence level; ** 5% confidence level and *10% confidence level.

Notes: To test for robustness besides including year dummies and varying firm size measures we also changed the variable *number of years rated in a row* to a dummy on the long-term relationship (3 years and 5 years), which didn't significantly change our results.

When comparing the estimation results for the two agencies, we see that the only statistically significant differences are in the effects of the number of years a product has been rated (negative and significant for Stiftung Warentest but positive and not significant for Morgen & Morgen). The effect of an insurer being a public organization is negative and significant for Morgen & Morgen but positive and significant for Finanztest. These differences could arise from the rating system differences (e.g. inclusion of historical data on operations by Morgen & Morgen) or simply the public rating agencies' preference for public institutions. The lack of a negative effect of public organizational form, for example, may reflect a desire to temper negative ratings in order to keep insurers in the sample over time.

Empirical Analysis of H4

Variables

To investigate Hypothesis 4a and 4b – bias in ratings due to competition between rating agencies – we use the full sample of rating data for both Morgen & Morgen and Stiftung Warentest. We do not employ insurance company data in these tests. We examine both downgrades and upgrades by the indicator variables $Lead\_HR_{t-j}$ and $Follow\_HR_{t-j}$, which express whether the regarded rating changes took place from a high (and therefore profit relevant) level of absolute rating as compared to lower (and therefore less profit relevant) rating categories as a basis. To be more precise, this indicator takes the value of 1 if the absolute rating from which the upgrade or downgrade was taken was of category 4 or 5 and it takes the value 0 whenever the rating change was based on an absolute rating of

category 3 or below.[21] As insurance companies use rating seals for advertising, they normally buy seals only for good ratings. Lower than category 4 rating seals are typically not used in advertising. Corresponding to the upgrade and downgrade variables, lags of these high rating indicator variables are used.

These high rating indicators should enable controlling for significant differences in reactions of the following rating agency on earlier published competitor ratings. Such differences could represent varying incentive levels for strategic rating. For example, to stay with the highest rating when the competitor issues a downgrade might be more effective as compared with sticking to a rating in category three when the competitor downgrades to two.

Using this logic, comparing the effects for high and low rating categories allows assessing whether an agency's reactions to competitor ratings seem more like strategic responses, in contrast to adjustments following real product enhancements or quality declines that are recognized by both agencies.

Methodology

We estimate two probit models to test the hypotheses. We assume that competitive pressure can be observed in strategic reactions of one rating agency to earlier published ratings of the other agency. Depending on whether the earlier publishing rating agency assigns an upgrade or downgrade to a particular product, the following agency could adjust its own rating decision in line with its profit maximizing objective as expressed in Hypotheses 4a and 4b. The model is built on the idea of Beaver et al. (2006) where they apply a so called Granger causality test (Granger, 1969) in order to examine whether credit rating changes issued by one agency can contribute to predict the changes in ratings issued by another agency. In accordance with our hypotheses we similarly test whether rating changes in product ratings depend on rating changes of competitor's ratings.

Consequently rating upgrades and downgrades of the later publishing agency (following agency) for product i in period t serve as dependent variables within the models ($Follow\_Down_{it}$ respectively $Follow\_Up_{it}$). Possible outcomes are binary. For the downgrade case $Follow\_Down_{it}$ takes the value of 1 whenever a product i receives a lower rating in period t as compared to period $t-1$ by the agency issuing the later rating in period t. It takes the value of 0 if the rating for product i is higher or equal in period t as compared to period $t-1$ by the same agency (non-downgrade). The model for the downgrade case will be used for both agencies in the role of the following agency and is specified as follows:

[21] Percentage shares of products in more profit relevant rating categories (4 and 5) as well as in less profit relevant rating categories (3 and below) are quite similar between the two rating agencies. Morgen & Morgen assigns high ratings to 74.69 percent of the 478 repeatedly and commonly rated products over the years and 25.31 percent of them receive ratings in the categories 3 and below. Stiftung Warentest assigns high ratings to 78.45 percent of the same products and rate 21.55 percent of these products with 3 and below.

$$\Pr(Follow\_Down_{it} = 1)$$
$$= \Pr(\beta_0 + \sum_{j=1}^{T} \beta_1 Follow\_Down_{t-j} + \sum_{j=1}^{T} \beta_1 Follow\_HR_{t-j} + \sum_{j=1}^{T} \beta_1 Lead\_Down_{t-j}$$
$$+ \sum_{j=1}^{T} \beta_1 Lead\_HR_{i-j} + \epsilon_{it})$$

For the upgrade case the dependent variable $Follow\_Up_{it}$ takes the value of one if the considered product i receives a higher rating in period t as compared to period $t-1$ by the agency issuing the later rating in period t. It takes the value of zero for the non-upgrade case. It will again be used for both agencies in the role of the following agency and is specified as follows:

$$\Pr(Follow\_Up_{it} = 1)$$
$$= \Pr(\beta_0 + \sum_{j=1}^{T} \beta_1 Follow\_Up_{t-j} + \sum_{j=1}^{T} \beta_1 Follow\_HR_{t-j} + \sum_{j=1}^{T} \beta_1 Lead\_Up_{t-j}$$
$$+ \sum_{j=1}^{T} \beta_1 Lead\_HR_{t-j} + \epsilon_{it})$$

Clustered standard errors on the insurer level are applied to avoid bias resulting from firm characteristics. Each model is conducted under consideration of one past rating period concerning own and competitor ratings. For Stiftung Warentest, as the later publishing agency, we additionally consider own and competitor ratings published two periods before the current Finanztest rating. The justification for this is as follows: Morgen & Morgen usually publishes their ratings in April whereas Stiftung Warentest used to publish their Finanztest ratings in July. Based on this situation Finanztest might not be able to incorporate observations out of the April rating by Morgen & Morgen into their own July rating at such short notice. Therefore, we provide a second version of the probit model for Stiftung Warentest where a second lag of rating variables is taken into consideration. With regard to Morgen & Morgen they would most likely look at the last year's Finanztest rating since a sufficient amount of time lies between the ratings. Therefore, we outline results using one lag of rating variables for Morgen & Morgen as the following agency.[22]

[22] The Pseudo R^2 as outlined in Tables 4 shows as expected that considering Period $t-1$ helps to explain Finanztest rating changes. We also tested this for Morgen & Morgen ratings where this is not the case.

Estimation Results

There are 478 situations for which Stiftung Warentest might be in the position to decide whether to follow the Morgen & Morgen rating published in April of the same year. When the Morgen & Morgen rating of April of the previous year is used, there are 455 such situations. Morgen & Morgen itself faces 432 situations where they could adjust their ratings towards those of Stiftung Warentest.[23] Table 4 provides details on the distribution of observations with regard to upgrades and downgrades within the sample of products rated by both agencies. Using these observations as our database, we examine the influence of competitor downgrades and upgrades on each agency's own rating decisions.

Table 4: Up- and downgrades for commonly rated products

	Downgrades	Upgrades	No rating change	Total
Morgen & Morgen	53 (12.27)	33 (7.64)	346 (80.09)	432 (100.00)
Stiftung Warentest (when comparison based on MM_{t-1})	76 (15.90)	74 (15.48)	328 (68.62)	478 (100.00)
Stiftung Warentest (when comparison based on MM_{t-2})	73 (16.04)	71 (15.60)	311 (68.35)	455 (100.00)

Numbers in brackets indicate percentage shares.

Table 5 reports the results of the probit estimation for the two rating agencies Morgen & Morgen and Stiftung Warentest, whereby the agency outlined in the head row takes the role of the later publishing rating agency (follower). Hypothesis 4a states that in cases where the earlier publishing agency (leader) issues downgrades, profit maximizing goals incentivize the later publishing agency to avoid downgrades. Finding support for this hypothesis would consequently mean that the likelihood for downgrades by the follower is significantly negative related with downgrades of the other agency.

Although we can identify significant relationships between own downgrades and earlier published competitor downgrades for both agencies, we cannot confirm the hypothesis for either agency. Contrary to expectations, we find a positive impact of earlier published competitor downgrades on later published own downgrades. For the ratings of Morgen & Morgen we find a highly significant positive relationship between own downgrades in the current rating period and downgrades of Stiftung Warentest in the period before. Incorporating a second lag of rating variables into the model helps to explain Stiftung Warentest's rating downgrades: we find a significant influence of both Morgen & Morgen ratings of period t-1 as well as of period t-2, and the influence of downgrades in period t-1 are smaller compared with those of downgrades in period t-2. Again, however, we see a positive relationship instead of the hypothesized negative relationship.

[23] To be more precise, products must be rated by both agencies over at least two consecutive periods to be able to observe situations where both agencies need to decide on up- and downgrades simultaneously. Additionally, with regard to the earlier publishing agency rating data must be available for at least three periods in a row as the considered starting point in period t-1 is already a rating change instead of an absolute rating.

Table 5: Probit for Hypothesis 4a

Sample	Morgen & Morgen		Stiftung Warentest			
	N = 415		**N = 245**		**N = 153**	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
Period t-1						
Follower downgrade in period t-1	-0.190	0.548	-0.955	0.001***	-5.072	0.000***
Follower rating in category 4 or 5 in period t-1	0.160	0.553	0.884	0.008***	-3.883	0.000***
Leader downgrade in period t-1	0.922	0.000***	0.216	0.364	1.089	0.043**
Leader rating in category 4 or 5 in period t-1	0.254	0.434	-0.335	0.186	0.170	0.795
Period t-2						
Follower downgrade in period t-2					-0.134	0.597
Follower rating in category 4 or 5 in period t-2					5.206	0.000***
Leader downgrade in period t-2					0.503	0.092*
Leader rating in category 4 or 5 in period t-2					-0.349	0.586
Pseudo R^2	0.0905		0.0932		0.1994	

*** Indicates the difference from zero is statistically significant at the 1% confidence level; ** 5% confidence level and *10% confidence level.

Note: To test for robustness we also included year dummies with no change in results.

In addition, we observe no significant differences between the agencies' reactions to the competitor's downgrades on high rated products versus low rated products. The rating agencies appear to move their own ratings in the same direction as their competitor – for both high rated and low rated products.

These results are more consistent with actual declines in product quality or changing quality requirements leading to new ratings, rather than strategic changes in ratings due to raters' financial interests. Even if one interpreted one rater following the other in downgrading a product as a consequence of competition, the positive correlation in ratings downgrades would not cause customer harm in the manner that inflated product ratings might. The lack of differences in following behaviors for downgrades of high rated products and low rated products reinforces the impression that the rating agencies do not choose strategically whether to follow their competitor's downgrades.

Hypothesis 4b states that a rating agency will have strategic incentives to follow a product rating upgrade issued by the earlier publishing rating agency. Such reactions could help to prevent competitive advantages accruing to one agency that provides a rating upgrade. We again identify strategic reactions to the competitor's rating upgrades by looking for differential responses to a competitor's rating changes on high rated versus low rated products. Only when the rating achieved after the upgrade ranges in the categories 4 or 5 an insurance company would typically buy the rating seal. For this reason, a strategic response to the competitor's upgrade would be found in evidence that following a rating upgrade is more likely when the upgrade yields a high rating. To capture this, the

empirical models incorporate an indicator variable for whether a product was assigned a rating category of 3 or 4 before the upgrade.

Results of estimation, reported in Table 6, cannot confirm Hypothesis 4b. A competitor's previous upgrade is not a statistically significant covariate in the models of rating upgrades. This is true for both rating agencies, and holds for Stiftung Warentest in the two-period as well as the one-period lag setting. Additionally, there is no effect of the absolute level of the competitor's ratings: products rated high by the competitor are no more likely to receive an upgrade than products rated lower by the competitor. Taken together, these suggest that the rating agencies do not behave strategically by following competitor rating upgrades.

More interesting, even if not directly related to the hypothesis we study, we find that the absolute level of rating from which an upgrade is taken plays a role for both agencies. Products that would come into a saleable rating category after upgrading are significantly more likely to receive an upgrade than ratings ranging in the categories 1 or 2. This could be an indicator of upward bias in ratings driven by financial interests of the rating agencies in selling rating seals to insurers; however, it could also reflect heightened financial incentives of insurers to improve the quality of products just below the rating for which advertising via the seal would be profitable. Additional details regarding reasons for product upgrades would need to be obtained in order to identify which effect is at work.

Table 6: Probit for Hypothesis 4b

Sample	Morgen & Morgen		Stiftung Warentest			
	N = 415		**N = 209**		**N = 125**	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
Period t-1						
Follower upgrade in period t-1	-0.415	0.412	omitted		omitted	
Follower rating in category 3 or 4 in period t-1	1.421	0.000***	1.200	0.002***	1.800	0.016**
Leader upgrade in period t-1	-0.058	0.803	0.434	0.237	0.741	0.307
Leader rating in category 3 or 4 in period t-1	-0.380	0.149	-0.212	0.388	-0.113	0.892
Period t-2						
Follower upgrade in period t-2					-1.363	0.012**
Follower rating in category 3 or 4 in period t-2					-1.402	0.021**
Leader upgrade in period t-2					0.784	0.112
Leader rating in category 3 or 4 in period t-2					0.512	0.578
Pseudo R^2	0.1611		0.0797		0.1650	

*** Indicates the difference from zero is statistically significant at the 1% confidence level; ** 5% confidence level and *10% confidence level.

Note: To test for robustness we also included year dummies with no change in results.

Discussion and Policy Implications

This article has provided the first empirical analysis of product quality ratings in an insurance market. Based on panel data of ratings published by two German product rating agencies and data on the insurance companies providing the products for years 1999 to 2013, our estimates provide no evidence of systematic rating bias in this market and no evidence of strategic responses of raters' to changes in a competitor's product rating. This may be due to the fact that financial incentives and conflicts of interest that may lead to rating manipulation are less pronounced in this market than in some other markets for ratings. As reported in Meyr and Tennyson (2015), insurance product ratings are not commissioned by insurers, and the only cash flow between insurance companies and rating agencies are fees for rating seals used in insurers' advertising. Such fees are modest and represent a rather small share of rating agencies' total income.

Nevertheless, results of our empirical tests indicate some possible triggers for bias in insurance product ratings. Products of larger insurance companies receive higher ratings from both rating agencies, and public insurers receive significantly higher product ratings from the government rating agency. In addition, we find evidence that rating upgrades are more likely when the product will be assigned to one of the two highest rating categories as a result of an upgrade. Additional research into the reasons for these patterns would be helpful in determining the quality of information provided to consumers by the ratings.

To further evaluate the contribution of insurance product ratings to enhance the functioning of insurance markets, consumer awareness and the influence of ratings on the demand for insurance contracts should be taken into account in additional research. Research into the capability of insurance product ratings to enhance consumer decision making is also needed.

Since we do not have data on how consumers use the ratings, or data on other indicators of product quality beyond the ratings themselves, we cannot comment on the implications for consumers' decision making. Experimental data or data on consumer complaints may provide an avenue for further study of this important question.

Appendix A1

Variable	Obs	Mean	Std. Dev.	Min	Max
Age of insurance company	987	75.28	51.06	7	208
Log net premium (TEUR)	986	12.58	1.58	6.03	16.60
Number of rated products	989	3.59	2.86	1	22
Loss ratio (in %)	987	68.31	34.59	1	219
Cancellation quota (in %)	955	5.10	2.70	1	45
Average sum insured per contract (TEUR)	957	31.14	19.06	0.024	110.59

		Stock company	Mutual	Public-service enterprise	Establishment of foreign company
Organizational form of insurance company	3485	2,483 (71.23%)	809 (23.21%)	113 (3.24%)	8 (2.30%)

References

Akerlof, G., 1970, The Market for "Lemons": Quality Uncertainty and the Market Mechanism, *The Quarterly Journal of Economics*, 84(3): 599–617.

Assekurata-Studie zur Bedeutung von Ratings für den deutschen Versicherungsmarkt, 2006, Köln. http://www.assekurata.de/content.php?baseID=130&dataSetID=300.

Barney, J., 2001, Resource-Based Theories of Competitive Advantage: A Ten-Year Retrospective on the Resource-Based View, *Journal of Management*, 27(6): 643–650.

Beaver, W., C. Shakespeare, and M. Soliman, 2006, Differential Properties in the Ratings of Certified versus Non-Certified Bond-Rating Agencies, *Journal of Accounting and Economics*, 42(3): 303–334.

Berry-Stölzle, T., and P. Born, 2012, The Effect of Regulation on Insurance Pricing: The Case of Germany, *Journal of Risk and Insurance*, 79(1): 129–164.

Blume, M., F. Lim, and A. MacKinlay, 1998, The Declining Credit Quality of US Corporate Debt: Myth or Reality?, *The Journal of Finance*, 53(4): 1389–1413.

Bolton, P., X. Freixas, and J. Shapiro, 2012, The Credit Ratings Game, *Journal of Finance*, 67(1): 85-111.

Covitz, D., and P. Harrison, 2003, Testing Conflicts of Interest at Bond Rating Agencies with Market Anticipation: Evidence that Reputation Incentives Dominate: FEDS Working Paper No. 2003-68. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=512402.

Doherty, N., A. Kartasheva, and R. Phillips, 2012, Competition among Rating Agencies and its Impact on Information Disclosure, *Journal of Financial Economics*, 106(2): 308–330.

Dranove, D., and G. Jin, 2010, Quality Disclosure and Certification: Theory and Practice, *Journal of Economic Literature*, 48(4): 935–963.

Granger, C., 1969, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica: Journal of the Econometric Society*, 37(3): 424–438.

Leland, H., 1979, Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards, *The Journal of Political Economy*, 87(6): 1328–1346.

Lizzeri, A., 1999, Information Revelation and Certification Intermediaries, *The RAND Journal of Economics*, 30(2): 214–231.

Meyr, S., and S. Tennyson, 2015, Product Ratings as a Market Reaction to Deregulation: Evidence from the German Insurance Market, *Risk Management and Insurance Review*, 18(1): 77-100.

Poon, W., 2003, Are Unsolicited Credit Ratings Biased Downward?, *Journal of Banking & Finance*, 27(4): 593–614.

Pottier, S., and D. Sommer, 1999, Property-Liability Insurer Financial Strength Ratings: Differences Across Rating Agencies, *Journal of Risk and Insurance*, 66(4): 621–642.

Romeike, F., 2004, Rating von Versicherungsunternehmen, in: S. Hirschmann, and F. Romeike, eds., *Rating von Versicherungsunternehmen*, 5–48 (Cologne: Bank-Verlag).

Strier, F., 2008, Rating the Raters: Conflicts of Interest in the Credit Rating Firms, *Business and Society Review*, 113(4): 533–553.

Viscusi, W., 1978, A Note on "Lemons" Markets with Quality Certification, *The Bell Journal of Economics*, 9(1): 277–279.