
The International Review of Financial Consumers

www.eirfc.com

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

Patricia A. Born¹ · Stephanie Müller² · Sharon Tennyson³

ABSTRACT

Instruments such as product ratings can help to overcome information asymmetries in retail financial markets. However, the capacity of ratings to promote market transparency and consumer awareness depends critically on whether they are credible. This article provides an empirical investigation of insurance product ratings in Germany, with an emphasis on the potential sources of bias that could undermine rating credibility. The analysis employs a panel dataset containing ratings for disability insurance products from two rating agencies over a 15-year period. Using the existing literature 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. Our results suggest no major concerns regarding the credibility of insurance product ratings.

Keywords: product ratings, insurance, ratings bias

1. Introduction

Insurance products, especially those for life, health and disability coverages, are widely recognized for their complexity, and the difficulty of judging product quality is a central information problem facing consumers in these markets. 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 transparency concerns, 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.

The idea that quality certifications (e.g. ratings) by information intermediaries may remedy information asymmetries has a long history in the economics literature (Viscusi, 1978; Leland, 1979). The net welfare effects of adding quality certifications to a market depend critically, however, on the characteristics of the certifications themselves. In a comprehensive 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. A large set of market characteristics can lead to incentives for rating bias and imprecision, and the mere existence of a ratings market does not assure informational efficiency.

1 Florida State University

2 Ludwig Maximilian University

3 Dept of Policy Analysis and Management, Cornell University

An 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, financial incentives are structured to reward upwardly biased ratings. Using data from the market for insurance product ratings in Germany, this study examines whether upward rating bias appears to exist for insurance product ratings in that country.¹

The study contributes to the literature in two main ways. First, it provides new evidence on the validity of product ratings, albeit in a specific context. Second, 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.² If an information market can provide valid, unbiased ratings of products, the need for government intervention in the form of strict product regulation may be reduced.

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 nature of the product ratings and describe the rating market structure. In section III we describe the dataset on product ratings and the supplemental data used to complete our analysis. We then develop testable hypotheses by drawing on previous literature and the institutional features of the rating market in section IV, discuss empirical methodology in section V, and present results in section VI. A final section discusses our findings and provides policy implications.

II. Background

A. The Product Rating Market

Insurance product ratings are external assessments of

the quality of a specific 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 often weight 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 of the insurer.³

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 several European insurance markets.⁴ Ratings have become an important fixture in the German market. Rating seals that 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 the seals to identify products they prefer to sell as well as to justify their advice. Additionally, product ratings are published in consumer magazines and in online product comparisons. According to Romeike (2004), German consumers are very likely to consult ratings before choosing an insurance company (72%). A survey by Assekurata (2006) suggests that more than 80% of German consumers at least occasionally consult product ratings when they search for information on insurance products. Hülsken (2010) found that over 80% of sales intermediaries use product ratings as a basis for advice they give on life insurance and health insurance purchases.

Insurance product ratings in Germany are provided by several private agencies as well as by one government foundation.⁵ Private and public agencies show significant differences in objectives, groups targeted by the ratings, and revenue sources. While private agencies' primary goal is profit maximization, more consumer-orientated objectives direct the actions of the government agency.⁶

¹ See Meyr and Tennyson (2015) for more details on the development and operation of this market.

² 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.

³ See for example the "Guide to Best's Financial Strength Ratings".

⁴ See Berry-Stölzle and Born (2012) for a description of the deregulation in Germany.

⁵ Rating agencies providing insurance product ratings are not affected by the European regulation of rating agencies.

Government raters have no financial incentives to produce upwardly biased ratings. However, the lack of financial incentives may reduce the raters' effort and thus the quality of their ratings (Dranove and Jin, 2010). Berger et al. (2000) found that in markets where both government raters and private raters operate—for example, in banking—there is exchange of information between the two groups and this may improve the accuracy of their ratings.

In contrast to credit ratings, insurance product ratings in Germany are not commissioned or paid for by the insurance companies. Most rating agencies assess the products' quality on their own initiative and choice, and many product ratings use only publicly available information (so-called PI-ratings). 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. The use of product ratings in insurance brokerage is not required by law (in contrast to credit ratings or bond ratings, which need to be consulted for specific investment decisions). Thus, insurance product rating agencies depend greatly on customers' and brokers' awareness.

B. Ratings for Occupational Disability Products

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

We base our analysis 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. Profit-maximizing motives for upward bias in product ratings should be relevant only for the for-profit rating agency (Morgen & Morgen), and not for the government agency (Stiftung Warentest). Thus, we compare the ratings of the two agencies to look for differences that may indicate upward bias in for-profit 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 deliver the data, and thus the rating is effectively interactive. Morgen & Morgen has provided product ratings for most occupational disability products in all years. Stiftung Warentest has not maintained a consistent approach over time with their Finanztest ratings. In some years the Finanztest ratings focused on particular aspects or target groups of occupational disability insurance products; in other years, the foundation decided to rate products that could provide alternative solutions to cover disability risks. For Stiftung Warentest we therefore restrict our analyses to years in which they rated occupational disability products and targeted the majority of consumers, and we drop years with special focuses (2002, 2012).

C. 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 ratings are issued for 873 distinct occupational disability products. The sample of observations is larger for Morgen & Morgen because the company rates all products in the market in every year, while Finanztest typically rates only a subset of

⁶ 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...".

Table 1. Sample Characteristics: Rating Data

No. of Products	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

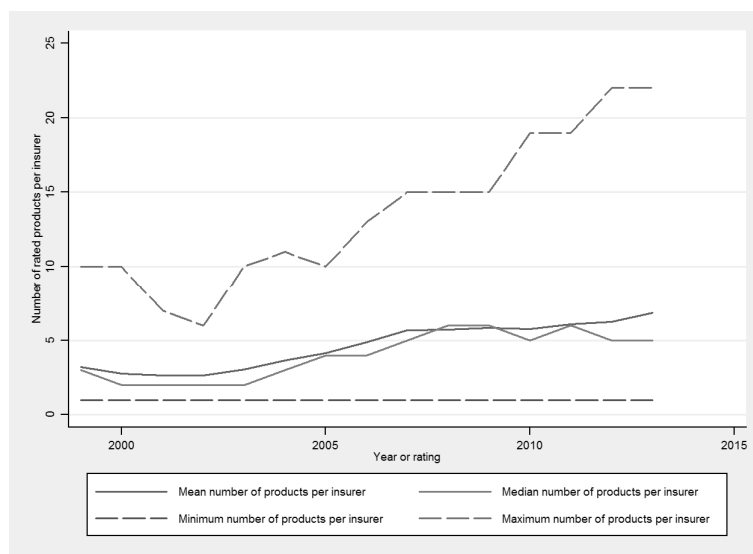


Figure 1. Number of Rated Products per Insurer and Year

products. The number of products rated in each year, for each agency, is shown in Table 1.

Figure 1 displays the mean, median, minimum and maximum numbers of products rated by at least one agency on the insurer-year-level. Since Morgen & Morgen aims to provide a comprehensive reflection of the considered market, the data provides evidence of an increasing number of products per insurer over time. This suggests an increase in complexity of the market over time.

Figure 2 shows the distribution of rating values published by Morgen & Morgen for our sample period. Morgen & Morgen issues ratings on a five-point-scale in which higher values indicate higher 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 revised its rating system in 2003, with the consequence that the proportion of these highest ratings is less than 40 percent after that.

Similarly, Figure 3 shows the distribution of product

ratings by Finanztest for the period 2000 to 2013. 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 increasing scale as used by Morgen & Morgen, with “defective” denoted by 1 and “very good” denoted by 5.⁷ 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.

⁷ 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”).

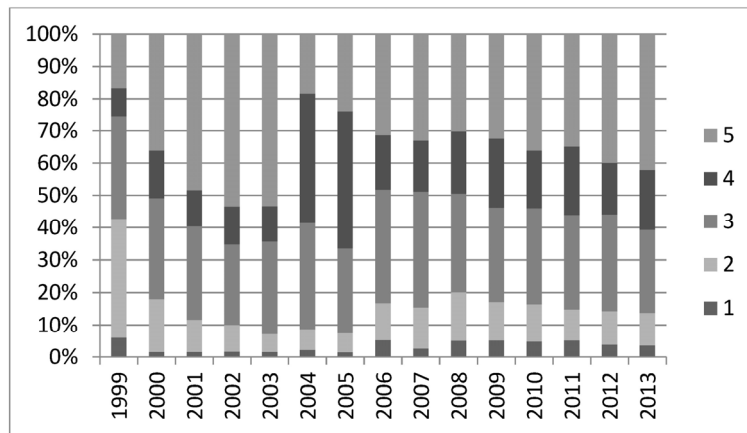


Figure 2. Distribution of Morgen & Morgen Ratings, 1999-2013

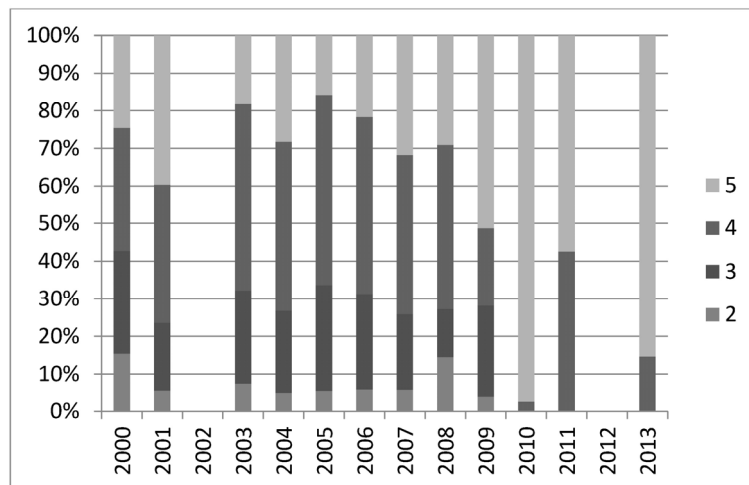


Figure 3. Distribution of Finanztest Ratings, 2000-2013

III. 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.⁸ Covitz and Harrison (2003) argue, however, that competition in rating markets will reduce the potential for 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 the quality of their ratings 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. For these reasons the insurance product rating market may have weak reputational penalties for inaccurate ratings.

Insurers do not typically commission product ratings, and rating agencies generally choose which products to rate. As discussed by Meyr and Tennyson (2015), the main source of revenue for private insurance rating agencies is from the provision of databases and software solutions to brokers to assist them with product comparisons. These characteristics of the market reduce the potential for rating bias (Bolton, Freixas and Shapiro, 2012). Nonetheless, there remains a profit incentive for upward bias in product ratings because insurers are more likely to purchase the rating seals of more highly rated products. The incentive to maximize the number of rating seals sold might motivate rating agencies to adjust their ratings upward. This could lead to upward bias in ratings and to little incentive for rating agencies to correct the upward bias over time.

Other considerations lead to predictions about specific patterns of potential ratings bias. First, insurers that offer a larger product portfolio might receive higher ratings as they could potentially buy a greater number of rating seals. In addition, larger insurance companies are usually

more familiar to customers and brokers and can therefore make a greater contribution to increasing a rating agency's prominence by the use of their rating seals in marketing. The desire to attract these larger insurers and keep them as customers could provide particular incentives to inflate ratings for products offered by larger insurers.⁹ Finally, long-term relationships may lead to ratings bias. 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.

Competitive dynamics in the product ratings market could also lead to upward bias. The desire to maintain relationships with insurers or to maximize the number of rating seals sold might motivate rating agencies to adjust their ratings in response to a competitor's assessment. Specifically, agencies may have profit incentives to follow a competitor's upgrade of a product rating but may have profit incentives not to follow a competitor's downgrade of a product rating. Maintaining a higher rating by not following a downgrade - especially if the product currently receives one of the highest ratings - will increase the likelihood of the rating seal being purchased and of receiving cooperation with future ratings. Similarly, following a competitor's upgrade will prevent the competitor from gaining a competitive advantage in rating seal purchases or future cooperation - especially if the upgrade moves the product into one of the highest ratings.

Based on the above reasoning, we propose several related hypotheses regarding bias in insurance product ratings.

- H1:** For-profit rating agencies will provide higher product ratings than a government rating agency.
- H2:** Product ratings increase with the number of products provided by an insurance company, all other factors held constant.
- H3:** Larger insurance companies receive higher product ratings, all other factors held constant.
- H4:** Product ratings increase with the number of years a product has been rated by an agency, all other factors held constant.

⁸ Lower ratings for unsolicited quality assessments could also result from sample selection bias or the rating procedure applied. Poon's (2003) results still hold when controlling for financial factors characterizing the rated companies.

⁹ 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.

H5a: Ratings downgrades from agency A are not followed by downgrades from agency B.

H5b: Ratings upgrades from agency A are followed by upgrades from agency B.

IV. Empirical Analysis of H1

Table 2 provides summary data on the relationship

between ratings provided by Morgen & Morgen and Stiftung Warentest, and includes only products rated by both agencies. Morgen & Morgen generally publishes their ratings in April whereas Finanztest ratings are normally published in July. The first panel of the table 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 nine months earlier. The table categorizes the relative ratings into groups, as follows. For each agency and product, we catalog whether the rating is less than

Table 2. Differences between Morgen & Morgen and Finanztest Ratings

Panel A: Comparison of MM ratings with 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 (0.0)	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

Panel B: Comparison of 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 parentheses indicate percentage shares.

the competitor rating by 4, 3, 2, or 1; whether both agency ratings are equal; and whether the agency rating is greater than the competitor rating by 1, 2, 3, or 4. The table displays the number of product ratings that fall into each grouping for each year and for the sample period as a whole, and shows in parentheses below this number the percent of ratings that fall in the grouping in parentheses. To conserve space, the table displays only the comparison groups for which there is a non-zero entry in at least one year of the sample period.

The data reveal that for the sample period as a whole, the majority of ratings provided by each agency are the same as the ratings published by the other agency. Specifically, 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.

However, looking at individual years in the sample period, 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, the Finanztest ratings appear increasingly to match those of Morgen & Morgen over time. These patterns fail to support the hypothesis that profit incentives lead to upward bias in Morgen & Morgen ratings and are more consistent with the Berger et al. (2000) observation that government raters may rely on information from private raters.

V. Empirical Analysis of H2-H4

A. Data and Variables

This analysis uses the panel data set of product rating data combined with insurance company data for the German life insurance market. We match each observed product rating in our data with company-specific information on the insurer that provides the rated product. Insurer data are obtained from Bisnode, a private provider of data on financial company characteristics and financial performance measures in Germany. Our resulting dataset includes company-specific data for 141 German life insurance companies in all years of our rating sample period. We are unable to obtain financial data for all insurers that offer disability insurance products, and as a result the merged dataset with rating data and insurance company data includes 3,383 ratings observations by Morgen & Morgen and 802 ratings observations by Stiftung Warentest.

Data on insurance companies include 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. Table 3 presents summary statistics for the insurance company variables included in our analysis.

Several of the variables are used to provide key information about the effects of provider-specific characteristics on product ratings. The insurer's size is characterized by the natural log of *net premiums* in a year.¹⁰ The size of the insurer's product portfolio is captured by the *number of rated products* in a year. This variable is specific to each rating agency. The duration of the agency-insurer business relationship is measured by the *number of years rated in a row*. This variable counts the number of consecutive ratings by an agency for a particular product before the current rating.¹¹

In addition to these three key variables, our models incorporate several other insurer characteristics as controls. We expect that older companies are more likely to provide

¹⁰ Results do not change when alternative measures of company size (total assets and equity capital) are used.

¹¹ Results do not change when we instead include an indicator of a long-term relationship. We tested alternative definitions of a long-term relationship as three years of rating in a row or five years of rating in a row, with similar results.

Table 3. Summary Statistics for Insurance Company Variables

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%)

Notes: The table reports descriptive statistics on insurance company data contained in the 1999-2013 panel data-set. All variables are indicated on the insurer-year-level except of data on the insurers' organizational form, which is presented on the product-year-level.

well-known brands and may have advantages in product design and distribution. The variable *company age* is included to control for these influences arising from a company's experience and establishment. We also consider the legal form of the insurance company which could be stock, mutual, public-service enterprise, or a foreign insurance company.¹² Special characteristics of the different legal forms regarding target groups, financing or decision-making might influence product design. As just one example, mutual insurers are owned by policyholders and could be expected to design products of better quality compared to stock companies. We include three indicator variables - *public*, *mutual* and *foreign* - in our models, with stock companies serving as the reference group.

Additionally, we include several measures of operating results as control variables. The *change in loss ratio* measures the annual change in the ratio of incurred losses to earned premiums. We expect a positive relationship between this variable and product ratings, since an increasing loss ratio indicates that policyholders are receiving an increasing level of claims payments per dollar of premiums paid. The *cancellation rate* of policies for each insurer in a year provides an indication of general consumer satisfaction with the insurer's products. We expect that product ratings are negatively associated with cancellation rates. Finally, the model includes the *average sum insured per contract* for an insurer in a year. Insurance companies with

larger exposures may be expected to have a clientele that demands higher product quality and therefore receive higher quality ratings than companies with smaller exposures.

B. Methodology

For our empirical analysis, we utilize an ordered multinomial model using data for each product and year.¹³ The rating outcome for product *i* in year *t*, denoted by R_{it} , will serve as the 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 μ_1 to μ_5 represent unknown thresholds for the observed rating categories:

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

We estimate the model using an ordered probit regression that examines the influence on the rating outcome of insurance company size, number of products in its portfolio and length of business relationship, after accounting for effects of the control variables. Standard errors are clustered at the insurer level to account for within-firm

¹² According to the national Insurance Supervision Act insurance companies in Germany are restricted to these four organizational forms. Besides stock and mutual insurers, companies established under public law have a long tradition as insurance providers in Germany.

¹³ The model design is based on Blume et al. (1998).

Table 4. Ordered Probit Estimates for Hypotheses 2 to 4

Full sample	Morgen & Morgen				Finanztest			
	Model IA:		Model IB:		Model IA:		Model IB:	
	coefficient	p-value	coefficient	p-value	coefficient	p-value	coefficient	p-value
	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	

Notes: Standard errors are clustered at the insurer level. *** Indicates the difference from zero is statistically significant at the 1% confidence level; ** 5% confidence level and *10% confidence level.

correlation over time. To control for unobservable year-specific effects in the 15 years of our sample, some versions of the estimates include year dummies.¹⁴ The ordered probit model is described by the following equation:

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

Since information on the “true” quality of products is not available, interpreting the effects per rating agency alone cannot provide sufficient information as to whether there is a rating bias or not. Comparing the effects of the two agencies on the basis of direction and significance¹⁵, however, makes it possible to use such interpretations as measures of a potential rating bias. These interpretations may also be informed by the fact that the regression analysis is based on one private and one governmental rating agency, each of which may be influenced by different incentives.

¹⁴ 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.

¹⁵ In the case of the ordered probit model applied here a direct comparison of the magnitude of the respective effects is not possible because of differing variances in the latent variable μ_{it} between models.

C. Estimation Results

Table 4 reports the results of the ordered probit estimation for the two rating agencies, Morgen & Morgen and Stiftung Warentest, separately. The estimates show that the effect of company size on a product rating is significant and positive for both rating agencies. 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 negative and significant for Stiftung Warentest ratings, and positive but not statistically significant for Morgen & Morgen ratings. These results provide distinctly mixed evidence with respect to our hypotheses regarding the insurance company features that could produce upwardly biased ratings due to the rating agency’s desire 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. However, there is no difference in the direction or statistical significance of these effects between the private agency and the government agency, so there is no strong indication of a rating bias. This result seems more consistent with the literature on industrial organ-

Table 5. 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)	76 (15.90)	74 (15.48)	328 (68.62)	478 (100.00)
Stiftung Warentest (when comparison based on MM _{t-1})	73 (16.04)	71 (15.60)	311 (68.35)	455 (100.00)

Notes: Numbers in parentheses indicate percentage shares.

ization and strategic management, which suggest that larger firm size may reflect higher productivity or other competitive advantages (e.g., Barney 2001). That is, it is possible that larger firms simply offer better products.

Examining estimated effects of the control variables, we see that for both rating agencies, insurance company age and average contract size are positive and statistically significant. Although less robust in statistical significance, products offered by subsidiaries of foreign insurers also receive higher ratings. These results suggest that products offered by larger and more-established insurers tend to be higher quality. Products offered by mutual insurers receive significantly lower ratings from Stiftung Warentest but there is no statistically significant effect of ownership form on Morgen & Morgen ratings. Interestingly, the effect of an insurer being a public-service enterprise is negative and statistically significant in Morgen & Morgen ratings, but positive and statistically significant in Stiftung Warentest ratings. Stiftung Warentest ratings are also sensitive to the change in loss ratio, showing the hypothesized effect that increases in the loss ratio lead to higher product ratings. These findings suggest rating system differences under which the public rating agency may favor public institutions and products with lower expense charges.

VI. Empirical Analysis of H5

A. Data and Variables

To investigate whether raters adjust product ratings in response to a competitor's changes, we begin with the sample of products that are rated by both Morgen & Morgen and Stiftung Warentest. To be included in

this analysis a product must be rated by both agencies over at least two consecutive periods, so that we are able to observe situations where both agencies decide on upgrades and downgrades. Additionally, data must be available for at least three periods in a row for the earlier publishing agency, because the comparison point in period $t-1$ is a rating change instead of an absolute rating.

In this analysis, we make use of differences in the two agencies' ratings publication dates. As noted previously, Morgen & Morgen usually publishes their ratings in April while Stiftung Warentest publishes their Finanztest ratings in July. Morgen & Morgen is the following agency in April of year t compared with Stiftung Warentest's leading ratings published in July of year $t-1$. Stiftung Warentest is the following agency in July of year t compared with Morgen & Morgen's leading ratings published in April of year t .¹⁶ There are 478 situations in our dataset for which Stiftung Warentest might be in the position to decide whether to follow a change in the Morgen & Morgen rating published only some months before (t) and 455 situations where Stiftung Warentest could decide whether to follow a rating change by Morgen & Morgen made in April of the preceding year ($t-1$). Morgen & Morgen faces 432 situations in which they could adjust their rating in response to a ratings change by Stiftung Warentest. Table 5 provides details on the distribution of observations with regard to upgrades and downgrades within the sample of products rated by both agencies.

¹⁶ It seems possible that Stiftung Warentest might not be able to incorporate observations of the April Morgen & Morgen rating into their own July rating, due to the short notice. For this reason we also estimated models with Stiftung Warentest as the "following" agency in July of year t that include Morgen & Morgen's "leader" ratings in April of year $t-1$. This reduces the number of observations available, but inferences remain similar.

B. Methodology

Using the observations in Table 5 as our database, we examine the influence of competitor downgrades and upgrades on each agency's own rating decisions. We estimate two probit models, one for ratings downgrades and one for ratings upgrades. In accordance with our hypotheses, we test whether changes in product ratings depend on previous rating changes of the competitor.¹⁷ Each model is estimated separately for each rating agency in the role of the following agency. We do not include insurance company control variables in these models. Standard errors are clustered at the insurer level to account for within-insurer correlation in ratings changes across products.¹⁸

We denote rating upgrades and downgrades of the later-publishing agency by agency B in the empirical models. The change rating for product i in period t serve as dependent variables in our estimated models (B_Down_{it} respectively B_Up_{it}). Possible outcomes are binary. B_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 (agency B) 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). B_Up_{it} takes the value of 1 whenever a product i receives a higher rating in period t as compared to period $t-1$ by the agency (B) issuing the later rating in period t . It takes the value of 0 if the rating for product i is lower or equal in period t as compared to period $t-1$ by the same agency (non-upgrade).

Our models also include an indicator for the level of the product rating before the change. As noted previously, insurance companies use rating seals for advertising but normally buy only the seals for good ratings. A product rating lower than 4 is typically not used in advertising. Thus, for rating downgrades a change of rating from a starting point of 4 or 5 is more consequential (negatively so) than a downgrade from a starting point of 3 or below. For rating upgrades, a change of rating

from a starting point of 3 or 4 is more consequential (positively so) than a change from any other starting point. Thus, indicators for these critical starting points provide additional evidence of 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.¹⁹

The model specification for rating downgrades is shown in the equation below. B_Down_{it-1} is an indicator of a rating downgrade (or not) of product i in the previous year by the following firm, and B_45_{it-1} is an indicator of product i receiving a rating of 4 or above in the previous year from this same firm. A_Down_{it-1} is an indicator of a rating downgrade (or not) of product i in the previous year by the leader firm (agency A), and A_45_{it-1} is an indicator of product i receiving a rating of 4 or above in the previous year from the leader firm. If raters are changing their ratings strategically, we expect to observe a smaller likelihood of downgrades from high ratings ($\beta_1 < 0$) and no follower response to leader downgrades ($\beta_2 = 0$).

$$\Pr(B_Down_{it} = 1) = \Pr(\alpha_i + \beta_1 B_{45_{it-1}} + \beta_2 A_Down_{it-1} + \beta_3 A_{45_{it-1}} + \epsilon_{it})$$

The estimation model for product rating upgrades is constructed analogously. B_Up_{it} and A_Up_{it} take the value of one if the considered product i receives a higher rating in period t than in period $t-1$, by the follower firm (agency B) and the leader firm (agency A), respectively: B_34_{it-1} is an indicator of product i receiving a rating of 3 or 4 in the previous year from the following agency, and A_34_{it-1} is an indicator of product i receiving a rating of 3 or 4 in the previous year from the leader firm. If raters are changing their ratings strategically, we expect to observe a higher likelihood of upgrades from midlevel ratings ($\beta_1 > 0$) and a positive follower response to leader upgrades ($\beta_3 > 0$).

$$\Pr(B_Up_{it} = 1) = \Pr(\alpha_i + \beta_1 B_{34_{it-1}} + \beta_2 A_Up_{it-1} + \beta_3 A_{34_{it-1}} + \epsilon_{it})$$

¹⁷ 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.

¹⁸ To test for robustness, we also estimated models that include year dummies, with no change in results.

¹⁹ 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.

C. Estimation Results

Table 6 reports the results of the probit estimation for the downgrade models. The agency listed in the header row takes the role of the later-publishing rating agency (agency B). The estimates reveal no relationships that suggest strategic changes to product ratings. For 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. Stiftung Warentest’s rating downgrades are significantly influenced by Morgen & Morgen ratings of period $t-1$, and again we observe a positive relationship instead of the hypothesized negative relationship. For both agencies, counter to the predictions of the strategic rating hypothesis, highly-rated products are more likely to receive a downgrade. There are no statistically significant effects on downgrades of the competitor’s previous high or low rating for a product.

In sum, in downgrading product ratings, 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.

Table 7 reports the results of the probit estimation for the ratings upgrade models.

Results of estimation reveal that a competitor’s previous upgrade is not a statistically significant covariate in the models of rating upgrades. This is true for both rating agencies. Additionally, there is no effect of the absolute level of the competitor’s ratings: products with midlevel ratings 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 strategically follow their competitor’s rating upgrades. Interestingly, however, we do 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 private rating agency in selling rating seals to insurers. However, given that there are no significant differences between the private rating agency Morgen & Morgen and the non-profit organization Stiftung Warentest, these

Table 6. Probit Estimates for Follower Downgrades

	Morgen & Morgen		Stiftung Warentest	
	N = 432		N = 355	
	coefficient	p-value	coefficient	p-value
Downgrade by agency A in period t-1	0.855	0.000***	0.541	0.028**
Rating of agency B in category 4 or 5 in period t-1	0.638	0.057*	0.796	0.033**
Rating of agency A in category 4 or 5 in period t-1	-0.079	0.771	-0.120	0.702
Pseudo R ²	0.0944		0.0559	

Notes: Standard errors are clustered at the insurer level. *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

Table 7. Probit Estimates for Follower Upgrades

	Morgen & Morgen		Stiftung Warentest	
	N = 432		N = 355	
	coefficient	p-value	coefficient	p-value
Upgrade by agency A in period t-1	1.293	0.615	0.266	0.177
Rating of agency B in category 3 or 4 in period t-1	1.217	0.000***	1.276	0.000***
Rating of agency A in category 3 or 4 in period t-1	-0.230	0.389	-0.183	0.318
Pseudo R ²	0.1311		0.1220	

Notes: Standard errors are clustered at the insurer level. *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

results may also indicate that insurers have financial incentives to improve the quality of their products to a level just below that needed for a rating that would justify advertising via the seal. To identify which effect these results indicate, more data are necessary.

VII. Discussion and Policy Implications

Based on panel data of ratings published by two German product rating agencies and data on the insurance companies providing the rated products for years 1999 to 2013, our estimates provide no strong evidence of systematic rating bias. Nevertheless, our results show that products sold by larger insurance companies receive higher ratings. This finding could indicate that the ratings of products supplied by big players in the insurance market are biased upward, because this outcome is observed in the ratings of both the for-profit Morgen & Morgen agency and for the government agency, Stiftung Warentest. However, it is also possible that larger insurers simply offer higher quality products. Our analysis also did not yield any evidence that rating agencies respond strategically to changes in ratings published by their competitor. However, there is some evidence that an agency is more likely to upgrade a product when this product receives one of the two highest ratings as a result of an upgrade. Because insurers usually buy rating seals only their products that receive high ratings, this rating pattern could indicate upward bias in ratings. Again, however, our results show this pattern for both agencies, suggesting that the incentives may come from the insurers' side - to improve products that are marginally below the rating level for which the seal would have value.

Our results also show no strong divergence of product ratings between the private agency Morgen & Morgen and the government agency Stiftung Warentest. This, and the lack of evidence of upward bias, implies that insurance brokers and consumers in Germany could be reasonably confident that such ratings can be used in individually appropriate consulting situations and purchase decisions. This positive outlook may be due to the governance features of this market, in which insurance companies do not commission or pay for product ratings. The primary clients of the private product rating agencies are insurance

agents and brokers, who purchase software licenses and consulting support from the agencies.

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.

References

- Assekurata-Studie zur Bedeutung von Ratings für den deutschen Versicherungsmarkt (2006). Köln. <http://www.assekurata.de/content.php?baseID=130&dataSetID=300>.
- 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* Vol. 42, No. 3, pp 303-334.
- Berger, A. N., S. M. Davies, and M. J. Flannery (2000). "Comparing Market and Supervisory Assessments of Bank Performance: Who Knows What When?". *Journal of Money, Credit and Banking* Vol. 32, No. 3, pp 641-667
- Berry-Stölzle, T., and P. Born (2012). "The Effect of Regulation on Insurance Pricing: The Case of Germany". *Journal of Risk and Insurance* Vol. 79, No. 1, pp 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* Vol. 53, No. 4, pp 1389-1413.
- Bolton, P., X. Freixas, and J. Shapiro (2012). "The Credit Ratings Game". *Journal of Finance* Vol. 67, No. 1, pp 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.
- Dranove, D., and G. Jin (2010). "Quality Disclosure and Certification: Theory and Practice". *Journal of Economic Literature* Vol. 48, No. 4, pp 935-963.
- Granger, C., (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods". *Econometrica: Journal of the Econometric Society* Vol. 37, No. 3, pp 424-438.
- Hülksen, M., (2010). "Der Einsatz von Produktratings bei der Versicherungsvermittlung". *AssCompact* (December), pp 112-113.
- Leland, H., (1979). "Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards" *The Journal of Political Economy* Vol. 87, No. 6, pp 1328-1346.
- 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* Vol. 18, No. 1, pp 77-100.
- Poon, W., (2003). "Are Unsolicited Credit Ratings Biased Downward?". *Journal of Banking & Finance* Vol. 27, No. 4, pp 593-614.
- Pottier, S., and D. Sommer (1999). "Property-Liability Insurer Financial Strength Ratings: Differences Across Rating Agencies". *Journal of Risk and Insurance* Vol. 66, No. 4, pp 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* Vol. 113, No. 4, pp 533-553.
- Viscusi, W.K., (1978). "A Note on "Lemons" Markets with Quality Certification". *The Bell Journal of Economics* Vol. 9, No. 1, pp 277-279.

Received/	2020. 03. 30
Revised/	2020. 05. 25
Accepted/	2020. 06. 01