©*The Journal of Risk and Insurance*, 1999, Vol. 66, No. 4, 621-642

Property-Liability Insurer Financial Strength Ratings: Differences Across Rating Agencies

Steven W. Pottier
David W. Sommer

Abstract

Regulators, investors, consumers, and insurance brokers use insurer financial strength ratings to evaluate the insolvency risk of insurers. This article investigates the factors influencing the decision to obtain a rating or multiple ratings, the determinants of ratings for the three major insurer rating agencies, and reasons for differences in ratings across agencies. This study indicates that insurers obtain ratings to reduce *ex ante* uncertainty about insolvency risk. It also provides evidence that specific rating determinants and their weights differ across agencies. Evidence of sample selection bias is found only in relation to Best's ratings.

Introduction

Insurer financial strength ratings provide a rating agency's opinion of the insurer's overall financial strength and ability to meet its policyholder obligations.[1] As such, ratings are meant to be summary measures of insolvency risk. In this article, we examine the property-liability insurer financial strength ratings of the three major insurer rating agencies, A.M. Best, Moody's Investors Service and Standard and Poor's (S&P). Our purpose is to address three fundamental questions regarding insurer ratings. First, what are the factors that influence whether or not an insurer chooses to obtain a rating or to obtain multiple ratings? Second, what are the determinants of insurer ratings from the three major rating agencies? Finally, what factors help explain differences in ratings for the same insurers across agencies? In addressing these questions, we control for the potentially important issue of sample selection bias, which can result from estimating empirical models using only data on rated insurers when there are systematic reasons why some insurers have chosen to be rated and others have chosen not to be rated.

Steven W. Pottier and David W. Sommer are Assistant Professors of Risk Management and Insurance at the Terry College of Business, University of Georgia. The authors gratefully acknowledge the helpful comments of the Editors and two anonymous referees.

[1] The rating process involves both quantitative and qualitative evaluations of an insurer's financial condition and performance. For a thorough discussion of the methods and procedures of the rating agencies, see Klein (1992).

This research is motivated by a number of factors. First and foremost is the critical importance of insurer ratings. As discussed more fully below, insurer financial strength ratings are heavily relied upon by insurance agents, brokers, and consumers, are used by insurers in their advertising, provide a tool for regulators to assess insurer risk, and are often used in academic research as measures of insolvency risk. Interest in ratings has been particularly strong since the rate of property-liability insurer insolvencies began to increase rapidly in the late 1980s. Given their great importance, it is valuable to examine these ratings.

The first step in understanding ratings is analyzing the determinants of ratings. Rating determinants models help establish the relative importance given to various factors in the rating process. This can help insurer management predict what their rating might be before deciding to apply for it, as well as assess the impact of proposed actions on their existing rating. For example, it is known that ratings influence insurer investment policies (Jones, 1994).

It is also very important to understand why insurers choose to obtain or not obtain a rating from a particular agency or agencies. Unlike bond ratings, insurer financial strength ratings are not required by regulation. They are voluntary, and while the vast majority of insurers choose to obtain a rating from A.M. Best, only a fraction opt to apply for a second or third rating from Moody's and/or S&P. It is particularly interesting to explore the motivations of insurers in obtaining these additional ratings, especially given the fact that the ratings given by Moody's and S&P tend to be lower on average than the ratings given by Best.

Finally, it is very important to understand the differences in ratings across rating agencies, as many insurers receive ratings from two or three rating organizations that do not agree. As stated earlier, insurer ratings are used for a variety of important purposes. For example, some corporate purchasers of insurance set minimum rating criteria for potential insurers. In order to establish these criteria in a logical manner, users need to know whether or not a rating of, say, "adequate" by one agency is equivalent to a rating of "adequate" by another. It is also worthwhile to know whether differences in ratings across agencies appear to be random or can be systematically explained. Related to this, another important question is whether differences in average ratings are truly driven by differing rating models, or are simply the result of selection bias.

RELEVANT LITERATURE

The determinants of debt ratings have been examined in many studies (see Kaplan and Urwitz, 1979; Ederington, 1985). Ederington and Yawitz (1987) provide an excellent survey of earlier work. The explanatory variables used include measures of profitability, capitalization, interest coverage, liquidity, debt subordination status, and industry indicators. Ederington (1986) focuses on bonds with split ratings; i.e., bonds that receive different ratings from different rating agencies. In examining bond ratings from Moody's and Standard and Poor's, Ederington (1986) does not find any evidence that split ratings result from differences in rating standards (i.e., different cutoff points) or weights attached to rating determinants. He concludes that split ratings represent random differences of opinion. Moon and Stotsky (1993) examine municipal bond ratings and find that split ratings reflect differences in the weights attached to rating determinants and differences in cutoff points. They also find limited evidence of sample selection bias. Cantor and Packer (1997) examine corporate debt ratings and test whether observed rating differences reflect different rating models or simply result from sample selection bias. Since virtually all publicly-traded corporate debt is rated by both Moody's and Standard and Poor's, they consider what types of firms are likely to seek a third rating. They find limited evidence of sample selection bias. Their results suggest that observed differences in average ratings reflect differences in rating models.

The academic literature on the determinants of insurer financial strength ratings is very limited. Pottier (1997) examines the determinants of Best's life insurer ratings and finds casual evidence suggestive of selection bias, but does not perform any formal tests for it or control for it econometrically. In addition, insurer ratings have been widely used as measures of insolvency risk and financial quality (Adiel, 1996; Anthony and Petroni, 1997; Cummins and Danzon, 1997; Pottier, 1998). These studies have only used Best's ratings.

The current article is most similar to Cantor and Packer (1997). Aside from the fact that they analyze only rating differences and not rating determinants, the key differences between their study and this one lie in the nature of the ratings being studied. Cantor and Packer analyze ratings that assess the default probability of particular bond issues, whereas we analyze ratings that assess the claims-paying ability of entire entities. By nature, their ratings are on publicly-traded securities and are designed for use primarily by investors, whereas the ratings analyzed here are on insurers of all sizes and ownership structures and are designed for use primarily by insurance consumers rather than investors. As such, the motivations for obtaining ratings may be expected to differ between the two types of ratings. In addition, Cantor and Packer (1997) compare bond ratings from agencies that specialize in bond ratings. The agencies we compare include one (A.M. Best) which focuses exclusively on insurer financial strength ratings and two (Moody's and S&P) that focus primarily on bond ratings and only recently entered the field of insurer financial strength ratings.

This article makes significant contributions to the insurance literature. As stated above, while insurer financial strength ratings have long been recognized as important and used in academic research, very little attention has been paid to the determinants of ratings. No study of insurer ratings has examined the decision to obtain a rating or multiple ratings, or analyzed the potential impact of sample selection bias. In addition, this is the first study to investigate differences in insurer ratings across rating agencies.

The Role of Insurer Ratings

Insurance company ratings provided by private rating agencies are vitally important to investors, regulators, consumers, insurers, and insurance agents/brokers. Insurers use ratings in their advertising to assure buyers of the firm's strength. Insurance buyers use them in choosing their insurance companies and/or deciding how much they are willing to pay for insurance from particular firms. Brokers and agents often will not recommend coverage with unrated insurers or insurers with ratings below some threshold of financial strength (Moody's, 1998), and many corporate insurance buyers require that all their insurers be highly rated. Strong financial ratings give insurers better access to capital markets. Insurer ratings also have a direct impact on the cost of capital, since the primary source of debt capital to insurers is policy liabilities, and lower rated firms will likely have to sell their policies at lower prices compared to higher rated firms (Doherty and Tinic, 1981; Berger, Cummins, and Tennyson, 1992). Ratings also provide a valuable tool for regulators in assessing the financial strength of insurers (Schwartz, 1994). As noted earlier, academic studies often use ratings as summary measures of solvency risk.

It should be noted that despite similarities, insurer ratings are quite different from corporate bond ratings.[2] First, as noted earlier, financial strength ratings are entirely optional in that there are no regulatory requirements to obtain a rating and agencies will not issue a full rating unless requested to do so by the insurer.[3] In addition, a bond rating applies to a particular debt issue, whereas an insurer rating applies to the entity itself and assesses the overall claims-paying ability of the insurer, since policyholder obligations must be met before payments are made to any other creditors or shareholders. Insurer ratings are particularly complex in that, unlike bond issues, which have fixed payments that are to be made at fixed times, claims payments involve financial obligations that are uncertain in both timing and amount.

The purposes of insurer ratings are also very different. Corporate bond ratings are used almost exclusively by investors and regulators, while the primary users of insurer financial strength ratings are insurance consumers and the agents/brokers who market insurance to consumers.[4] This point is illustrated by the fact that the majority of insurers rated by A.M. Best are either mutuals or privately-held stock firms. Additionally, the fact that many insurers with rated debt still choose to obtain one or more insurer financial strength ratings, and that these ratings often differ from their debt ratings, again illustrates the fact that bond ratings and insurer financial strength ratings do not serve the same purpose or measure precisely the same risk.

One interesting practical distinction between bond ratings and insurer ratings is that the bond rating agencies that rate a smaller proportion of bonds tend to issue higher

[2] For a description of the relationship between bond ratings and insurer financial strength ratings from the perspective of the rating agencies, see Buckley (1997) and Moody's (1998).

[3] As noted in footnote 6, S&P does issue a type of purely quantitative rating even on insurers that do not apply for a rating.

[4] Despite the existence of state guaranty funds, consumers have reason to care about the financial strength of their insurer. Insurance guaranty fund coverage is incomplete, and wide variations in coverage exist from state to state. Some lines of insurance are not covered at all and all states limit the maximum recovery per policy (typically $300,000). Also, 22 states employ net worth criteria to determine an insured's eligibility for their guaranty funds which particularly limits coverage for many commercial policyholders. State insurance guaranty funds use a post-insolvency assessment mechanism, except for New York's guaranty fund (see Standard and Poor's, 1996). Sommer (1996) found that because of the incompleteness of these funds, riskier insurers receive a lower market price for the insurance they sell than safer insurers. Cummins and Danzon (1997) also find that price is positively related to financial quality. However, the existence of guaranty funds potentially limits the impact of ratings on firm value, to the extent that it shields some policyholders from the full impact of insolvencies.

ratings than the agencies that rate almost all bonds (Cantor and Packer, 1997), whereas the agencies (Moody's, S&P) that rate a smaller proportion of insurers tend to issue lower ratings than the agency (A.M. Best) that rates most insurers. Another distinction is that there appears to be a greater divergence of opinion among rating agencies regarding insurer financial strength ratings than bond ratings.[5]

Almost 90 percent of eligible property-liability insurers in 1995 applied for a rating from A.M. Best, while only about 18 percent applied for a rating from S&P and only 10 percent applied for a rating from Moody's.[6] Given the overwhelming proportion of insurers that receive Best's ratings, it appears that insurers consider obtaining at least one rating essential. There are some potential explanations for why A.M. Best appears to be the agency of choice for most insurers for a first rating rather than S&P or Moody's. The most important factor is probably the historic relationship between Best and the insurance industry. A.M. Best focuses exclusively on the insurance industry, and has been in the insurer rating business for over 90 years. Moody's and S&P, on the other hand, while long established in the bond rating business, just entered the insurer financial strength rating field in the 1980s.[7] In addition, while all the agencies use a sliding cost scale, the minimum charge to obtain a Best rating is only $1,000, whereas the minimum cost of a Moody's or S&P rating is around $15,000.[8]

Hypotheses

Our empirical work consists of two sets of models, one for rating determinants and one for differences in ratings across agencies. Each model consists of two stages and controls for potential sample selection bias. For the rating determinants models, the first stage is a probit regression modeling the decision to obtain a rating, while the second stage is an ordered probit regression modeling the determinants of the rating. For the rating differences models, the first stage is a probit regression modeling the decision to obtain a particular pair of ratings, while the second stage is an ordered probit regression modeling the difference between those ratings.[9]

---

[5] For example, in the study by Ederington (1986), only 13 percent of the corporate bonds rated by Moody's and S&P received different ratings. In our sample, 56 percent of the insurers rated by these same two agencies received different financial strength ratings.

[6] Standard and Poor's issues two types of ratings, "interactive" and "quantitative" ratings. Only interactive ratings are considered in this study since these ratings must be applied for by the insurer and involve the full rating process. Quantitative ratings are assigned at the discretion of S&P and involve only publicly available, quantitative information. Since firms with quantitative ratings have not chosen to receive a rating, they are categorized here as not rated. Other agencies also issue insurer ratings. Weiss ratings, like S&P quantitative ratings, are not applied for and are based solely on publicly available data. Duff & Phelps rates 97 of the firms in our sample. However, 94 of the ratings are in the range of AA to A-, leaving very little dispersion for statistical analysis.

[7] Unlike Moody's and Standard and Poor's, A.M. Best does not rate insurer debt.

[8] Since the rating process typically entails meetings with insurance company management and responding to information requests from rating analysts, the cost of obtaining a rating is potentially much higher than the rating agency's fee.

[9] The details of the sample selection methodology employed will be discussed later.

The Decision To Be Rated

For each of the models, the same set of variables is used in the first stage probit regression. These variables are hypothesized to influence the decision by the firm to obtain a rating or multiple ratings. According to the theory of financial intermediation, the principal role of credit rating agencies is the reduction of *ex ante* uncertainty or informational asymmetry about a firm's economic value and probability of financial distress (Ramakrishnan and Thakor, 1984; Millon and Thakor, 1985). Thus, the more likely investors, consumers, and regulators are to have different opinions about the true insolvency risk of an insurer, the greater the demand for and value of a financial strength rating. In addition to uncertainty about an insurer's default risk, other factors are also expected to affect the decision to seek a rating, as described below.

Numerous variables are included in the model to proxy for the level of uncertainty about the firm's risk. Cantor and Packer (1997) argue that relatively high levels of profitability and leverage may be associated with greater uncertainty and thus a higher probability of obtaining an optional rating.[10] Therefore, our model includes a leverage measure and a profitability measure. A measure of premium growth is also included. While strong premium growth may be very positive for the financial health of the firm, growth by property-liability insurers is sometimes due to lower underwriting standards or underpricing (Harrington and Danzon, 1994). Because the effect of growth on firm value is ambiguous, relatively high rates of growth might be associated with greater uncertainty. Higher levels of investment in common stocks and in bonds of below investment grade quality (junk bonds) would also be associated with greater uncertainty, while higher levels of investment in cash would be associated with lower uncertainty because stocks and junk bonds are riskier and more difficult to value than cash. Selling more lines of insurance and selling insurance in more states would also increase uncertainty.[11] Therefore, line of business and geographic diversification measures are included.

The impact of the extent of reinsurance use on uncertainty is ambiguous. On the one hand, reinsurance involves the transfer of risk away from the primary insurer to the reinsurer. Berger, Cummins, and Tennyson (1992) suggest that reinsurance is a source of capital and a risk-diversification device. Thus, reinsurance usage may be expected to reduce uncertainty. On the other hand, reinsurance only accomplishes its purpose to the extent that reinsurers fulfill their obligations. Therefore, the risk of a heavily reinsured company largely depends on the risk of its reinsurers. This complicates the assessment of the primary insurer's true risk level. Rating agencies explicitly state that they consider the financial strength of a company's reinsurers in assigning a rating (Best, 1996; S&P, 1996; Moody's, 1998). The expertise of insurer rating agencies in assessing the quality and appropriateness of reinsurance should increase the value of obtaining a rating for heavily reinsured firms. Thus, higher levels of reinsurance might increase the probability of obtaining a rating or multiple ratings.

[10] Cantor and Packer (1997) consider bond ratings from Moody's and S&P to be mandatory because these agencies rate virtually all publicly-traded debt, while they consider ratings from Duff & Phelps and Fitch to be optional.

[11] In the terminology of Mayers and Smith (1981), writing insurance in more lines and across more states involves greater managerial discretion, and hence, greater uncertainty.

Another variable included in the first stage regression is the percent of premiums written in so-called "long-tail" lines of business.[12] Long-tail lines of insurance are those lines of business in which there is often a long period of time between the insured event and the final claim payment by the insurer. A variable measuring the extent of business in long-tail lines is included for two reasons. First, as with the variables discussed above, long-tail lines are associated with greater levels of uncertainty (Sommer, 1996; Fung et al., 1998). Second, buyers of long-tail lines of insurance are likely to be even more concerned about the financial health of their insurer than buyers of short-tail lines, since the value of the insurance purchased is dependent on the insurer being solvent until all claims associated with the policy are settled. Both the uncertainty hypothesis and the demand of long-tail insurance buyers would lead to a predicted positive association between long-tail business and the probability of obtaining a rating or multiple ratings.

The percent of business in personal lines is also included in the first stage.[13] While all insurance consumers have reason to be concerned about the financial strength of their insurers, the structure of the guaranty fund system gives commercial insurance buyers even greater reason to be concerned than personal insurance buyers. Commercial insurance buyers are much less protected by guaranty funds than personal insurance buyers. Also, as a consequence of due diligence requirements, many corporate insurance buyers will not purchase insurance from unrated insurers. It would therefore be expected that commercial buyers would demand ratings to an even greater extent than personal buyers. Thus, we expect a negative association between the percent of business in personal lines and the probability of obtaining a rating or multiple ratings.

Whether or not an insurer is publicly traded is also hypothesized to be an important determinant in the decision to obtain a rating. Compared to a mutual insurer or a closely-held stock insurer, a publicly-traded stock insurer has an additional group interested in its financial strength, namely, outside investors. As stated earlier, ratings can improve an insurer's access to capital markets and reduce its cost of obtaining capital. Any security issued by an insurance company is subordinate to the obligations to policyholders, so a claim-paying ability rating provides valuable information to potential investors. Just as an investor in the subordinate debt of a corporation might want to know the rating of the senior debt of that corporation, investors in the securities of an insurer might want to know the claim-paying ability rating of the insurer. Moreover, Thompson and Vaz (1990) find that investors value the certification function of rating agencies and that bond issuers are generally better off if they obtain more than one bond rating.[14] In addition, since publicly-traded companies are more likely to issue rated debt, the marginal cost of obtaining a claims-paying ability rating is probably reduced after having already obtained ratings on their debt

[12] Long-tail lines are defined as in Sommer (1996) to include auto liability, other liability, farmowners/homeowners/commercial multiple peril, medical malpractice, workers compensation, ocean marine, aircraft, and boiler and machinery. Examples of short-tail lines are fire, automobile physical damage, and earthquake.

[13] Personal lines are defined to include private passenger auto liability, auto physical damage, and homeowners.

[14] Specifically, they find that bond yields tend to be lower for issuers that obtain multiple ratings compared to issuers that obtain only one rating.

from the same agencies. For the above reasons, we hypothesize that being publicly traded is associated with a higher probability of obtaining multiple ratings.

In their equation for the decision to obtain an optional rating, Cantor and Packer (1997) include the natural logarithm of long-term debt outstanding as a measure of the potential benefit of an additional rating. This variable is used because the major benefit of obtaining an additional rating is a lower cost of debt, which should accrue in direct proportion to the amount of debt issued. Since in our context ratings are for claim-paying ability on policies written rather than for long-term debt, we include the natural logarithm of direct premiums written in our model following the same rationale. We expect that the more premiums written, the greater the likelihood of obtaining a rating.

Rating Determinants

The second stage of the rating determinants models contains numerous variables intended to measure various aspects of the risk of the insurer. The variables included represent factors that previous theory has indicated are important in determining insurer insolvency risk. These include variables reflecting capitalization (Kahane et al., 1986; MacMinn and Witt, 1987; Cummins, 1988; Doherty, 1989), asset and liability risk (Kahane et al., 1986; MacMinn and Witt, 1987; Cummins, 1988), liquidity (Kahane et al., 1986), size (Cummins et al., 1995; Cummins and Sommer, 1996), growth (Harrington and Danzon, 1994), diversification (Sommer, 1996), reinsurance usage (Borch, 1974; Berger et al., 1992), and profitability (Kahane et al., 1986; MacMinn and Witt, 1987).[15] Many of the variables are similar to measures used in articles studying the determinants of bond ratings, while others are explicitly stated as determinants of insurer ratings by the rating agencies themselves. Specifically, the variables include measures of liquidity (CASH), investment risk (JUNK, STOCKS), use of reinsurance (REINSURANCE), size (LN(ASSET)), leverage (CAPASSET), growth (CHGNPW), profitability (ROA), percentage of business in long-tail lines (LONGTAIL), geographic diversification (STATES), and line-of-business diversification (LOBHERF). Each of these variables is defined in the note to Table 4.

Determinants of Rating Differences Across Agencies

Ederington (1986) has proposed three explanations for differences in ratings across agencies. The first is that agencies agree on the default risk of the firm but apply different cutoff points for their ratings; the second is that agencies include different factors in their rating models or apply different weights to these factors; and the third is that differences merely reflect random variations in judgment. The first two of these reasons both imply that differences in ratings across agencies might be explained by the same variables that are rating determinants for the agencies.[16] Therefore, the second stage of the rating differences models contains the same variables that are included in the second stage of the determinants models.

[15] For a collection of articles dealing with financial models of insurer solvency, see Cummins and Derrig, 1989.

[16] The third reason implies that there is no difference in average ratings and that no variables would successfully explain any differences, since subjective differences in judgment would show up in the error term of any model of rating differences.

Data and Sample

Our sample consists of 1678 individual property-liability insurers.[17] The primary data sources for ratings were Best's *Key Rating Guide*, Property-Casualty (PC) Edition; Moody's *PC Insurance Financial Strength Rating List*, and S&P's *Insurer Solvency Review*, PC Edition. The ratings were effective as of July, 1996, and are based on the financial statement year ending December 31, 1995.[18] Data for financial ratios and other firm-specific variables are from the National Association of Insurance Commissioners (NAIC) annual statement database. Publicly-traded insurers are identified using the Compact Disclosure database, which provides information on subsidiaries of publicly-traded insurers and insurance holding companies. In order to be included in our sample, an insurer must have financial data available on the NAIC database necessary to calculate the various explanatory variables. As shown in Table 1, 1,510, 296, and 170 property-liability insurers are included in the sample of insurers rated by Best, S&P and Moody's, respectively.[19]

The dependent variables in our rating determinants models are based on the actual rating categories assigned by Best, Moody's, and S&P to property-liability insurers. Rating categories are combined as shown in Table 1 to conform with the correspondence of the verbal descriptions provided by the three rating agencies (also see Garber, 1994).[20] The four-level rating categories are denoted BCAT4, MCAT4, and SPCAT4. Each rating category is assigned an ordered numerical value as shown in Table 1, where a higher value indicates a higher rating. Table 2 presents summary statistics on selected firm-specific variables. Insurers rated by Moody's or S&P are larger and licensed in more states than insurers rated by Best. While over 70 percent of insurers rated by Moody's and S&P are publicly traded, only 32 percent of insurers rated by Best are publicly traded.

---

[17] Data on individual insurers are used. As pointed out by an anonymous referee, Best often assigns group ratings. This may result in correlation in the disturbances across group members. As a robustness check, we ran our Best's rating determinants model using only insurers that received a stand-alone (not group) rating and our results were qualitatively similar. Thus, using individual insurer data does not appear to be a problem.

[18] Annual financial statement data is filed with the National Association of Insurance Commissioners in March each year, and ratings based on the annual data are released between April and June.

[19] Of the 1,510 insurers rated by Best, 94 received Financial Performance Ratings (FPRs). Although the rating process for letter ratings and FPRs is the same, FPRs are assigned to firms that are either too small or too young to qualify for a letter rating. In this study, FPRs are grouped with the letter ratings according to the correspondence given in Best (1996). If these FPR firms are categorized as "not rated" the results for the Best's rating model are qualitatively unchanged. Results for the difference models are totally unaffected, since none of the FPR firms are rated by Moody's or S&P.

[20] Our results do not appear sensitive to any particular categorization of the ratings. As robustness tests, we tried running the determinants models using a number of different categorizations, including simply ranking the thirteen rating categories of each agency with no grouping. In each case, the results were qualitatively unchanged. In addition, we tested differences in coefficients across agencies (similar to Table 6) using nine categories for each agency, with consolidation only for the lowest category. Again, results were qualitatively unchanged compared to using four categories. We focus on the four-category specification because of the close correspondence between the verbal interpretations of the ratings for Best and the other two agencies that exists at this level of aggregation (see Table 1).

Table 1
Insurer Financial Strength Rating Symbols and Rating Distributions[a]

	Best			S&P			Moody's		
Interpretation[b]	BCAT	Number	BCAT4	SPCAT	Number	SPCAT4	MCAT	Number	MCAT4
Superior/Superior/Exceptional	A++	105	4	AAA	54	4	Aaa	14	4
	A+	219	4						
Excellent/Excellent/Excellent	A	392	3	AA+	32	3	Aa1	20	3
	A–	372	3	AA	37	3	Aa2	22	3
				AA–	39	3	Aa3	11	3
Very good/Good/Good	B++	115	2	A+	54	2	A1	40	2
	B+	146	2	A	38	2	A2	22	2
				A–	24	2	A3	11	2
Adequate/Adequate/Adequate	B	62	1	BBB+	4	1	Baa1	14	1
	B–	53	1	BBB	14	1	Baa2	5	1
				BBB–	0	1	Baa3	10	1
Uncertain claims-paying ability[c]	C++	10	1	BB+	0	1	Ba1	1	1
	C+	19	1	BB	0	1	Ba2	0	1
	C	9	1	BB–	0	1	Ba3	0	1
	C–	5	1						
	D	3	1						
Totals		1,510			296			170	

[a] BCAT, SPCAT, and MCAT are the actual rating levels for Best, S&P, and Moody's, respectively. BCAT4, SPCAT4, and MCAT4 represent rating categories after ratings have been consolidated to four levels. As can be seen, the four levels have corresponding verbal interpretations across rating agencies.

[b] These are the descriptions used by the three major rating agencies (Best/S&P/Moody's).

[c] The descriptions used by the rating agencies are more difficult to compare below the "adequate" level. All insurers rated by more than one of the rating agencies had Best's ratings of very good (B++, B+) or higher, except for one company rated by both Best and Moody's. Consequently, grouping all ratings below adequate into one category only affects the rating difference for one firm.

TABLE 2
Selected Summary Statistics

	Total Sample	Best	S&P	Moody's
Number of firms	1,678	1,510	296	170
Means:				
Assets (000s)	$409,255	$450,237	$1,601,283	$2,354,251
Net premiums written (NPW) (000s)	$135,531	$149,068	$502,898	$751,383
Personal lines NPW / Total NPW	37.5%	38.4%	30.1%	33.3%
Long-tail NPW / Total NPW	65.6%	67.0%	73.1%	76.1%
Number of states licensed	16.3	17.6	33.0	39.8
Publicly-traded	29.7%	32.4%	72.0%	75.3%

Spearman rank correlations of rating categories based on assigned (original) rating categories and combined rating categories were calculated. The correlation between Moody's and S&P ratings based on the original rating categories and the four-level rating categorization is .92, while the correlation between Best and Moody's is .62; between Best and S&P it is .67. These correlations suggest that we would find more differences of opinion regarding insurer risk between Best and the other two agencies than between Moody's and S&P. Correlations, however, only measure the relative level of agreement between rating agencies and might not capture differences in average ratings.

The dependent variables in our rating differences models represent the rating differences reduced to three levels.[21] Rating differences are calculated by assigning ordered numerical values to ratings, where a higher number indicates a higher rating. In the case of differences between Best and the other two agencies, the rating difference is defined as BCAT4-SPCAT4 for S&P compared to Best, and BCAT4-MCAT4 for Moody's compared to Best. The sample of companies rated by both Best and S&P includes 295 firms, and the sample rated by both Best and Moody's consists of 169 firms. The three levels represent same, higher by one level, and higher by two levels, because hardly any insurers rated by both Best and one of the other two agencies receive a lower rating from Best (see Table 3).[22] Best's and Moody's assigned ratings (based on four-level categorization) differ for almost 82 percent of insurers, while Best's and S&P's assigned ratings differ for over 64 percent of insurers. The sample of insurers rated by both Moody's and S&P includes 140 companies. In the case of Moody's compared to S&P, the rating difference is defined as SPCAT-MCAT,[23] and all differences are combined into three categories representing lower, same, or higher rating. It should be noted that we were able to exploit a finer partitioning of the rating scale in calculating the rating difference between Moody's and S&P because of the closer correspondence in the rating categories of these two agencies. Also, if rat-

[21] This approach is similar to Cantor and Packer (1997) in the case of their trinomial ordered probit model of rating differences.

[22] As shown in Table 3, Best rated only 5 insurers lower than S&P and 2 lower than Moody's. These firms were included in the "same" category.

[23] SPCAT and MCAT were assigned numerical values ranging from 1 for BB+ and Ba1 to 11 for AAA and Aaa.

ing differences between Moody's and S&P were calculated based on the four-level rating categorization used in the rating determinants model (i.e., SPCAT4-MCAT4), then almost 83 percent of the 140 insurers rated by these two agencies would be in the same rating category. The finer partitioning results in only 44 percent (62 out of 140) of the insurers rated by Moody's and S&P in the same rating category. As Table 3 shows, because so few rating differences are more than one category higher or lower, classification of differences into three categories results in minimal grouping, while reducing the sample sparsity in some cells. The summary statistics presented in Table 3 show that on average Best's ratings are higher than both S&P and Moody's, and S&P ratings are higher than Moody's. As a univariate test of rating differences, we compared the distribution of ratings across the four categories (three pairwise comparisons using BCAT4, SPCAT4, and MCAT4) and find that each of the actual rating distributions is significantly different from the others at the .01 level of significance based on the chi-squared test statistics.

It is interesting that Best's ratings tend to be higher than Moody's and S&P ratings. In

TABLE 3
Rating Differences[a]

	Best vs. S&P	Best vs. Moody's	S&P vs. Moody's
Higher by four	0	0	8
Higher by three	0	0	4
Higher by two	26	43	2
Higher by one	164	95	39
Same	100	29	62
Lower by one	5	1	24
Lower by two	0	1	0
Lower by three	0	0	1
Total	295	169	140

[a] The rating differences between S&P and Moody's are based on the original rating categories, similar to Cantor and Packer (1997). The four-level categorization is used to calculate differences between Best and the other two rating agencies (see footnote 20 for the rationale for using the four-level categorization for rating differences models involving Best).

the area of bond ratings, Cantor and Packer (1997) consider Moody's and S&P to be "mandatory" rating agencies, and others to be "optional" agencies. For insurer financial strength ratings, Best would most properly be called the "mandatory" agency, while Moody's and S&P are "optional." However, whereas Cantor and Packer found that for bond ratings the optional agencies tended to issue higher ratings than the mandatory agencies, just the opposite appears true for insurer ratings.[24] Here, the optional agencies tend to issue lower ratings than the mandatory agency. Thus, many firms are volunteering to obtain optional ratings that are in many cases lower than

[24] Cantor and Packer (1997) do not include any analysis of differences between Moody's and S&P ratings. Ederington (1986) finds no significant differences in the determinants of Moody's and S&P corporate bond ratings. It is interesting that while the corporate bond rating models of Moody's and S&P may not differ, their insurer rating models do, as will be seen later.

their Best's rating. Clearly, there must be potential benefits to obtaining additional ratings even if the additional ratings may be lower than the firm's first rating.

Methodology

Corporate bond ratings and insurer ratings alike are inherently ordered (Ederington, 1985; Ederington, 1986; and Greene, 1997). Since ratings and rating differences are ordinal variables, we use ordered probit to estimate the regression for these variables.[25] The ordered probit model was developed by McKelvey and Zavoina (1975).

The ordered probit model is based on the following specification:

$$y^* = \boldsymbol{\beta} x_i + \varepsilon_i \tag{1}$$

$$y_i = \begin{cases} 0 \text{ if } y^* \leq \mu_0 = 0 \\ 1 \text{ if } \mu_0 < y^* \leq \mu_1 \\ 2 \text{ if } \mu_1 < y^* \leq \mu_2 \\ \vdots \\ J \text{ if } y^* > \mu_{j-1}, \\ \text{and} \\ 0 < \mu_1 < \mu_2 < \cdots < \mu_{j-1}. \end{cases}$$

The variable of theoretical interest, y^*, is a continuous index of risk and is unobserved. The observed rating categories are assumed to represent ordered partitionings of this continuous scale, where y_i is the observed rating category for firm i, $\boldsymbol{\beta}$ is a vector of coefficients, $\boldsymbol{x}_i$ is a vector of explanatory variables for firm i, ε_i is a standard normal random error, and the μ_js are threshold parameters. A higher value of a variable with a positive coefficient, β, indicates a greater probability of a higher rating. For four response levels (i.e., rating categories), that is $J + 1 = 4$,

$$\begin{aligned} P(y=0) &= \Phi(-\beta x_i) \\ P(y=1) &= \Phi(\mu_1 - \beta x_i) - \Phi(-\beta x_i) \\ P(y=2) &= \Phi(\mu_2 - \beta x_i) - \Phi(\mu_1 - \beta x_i) \\ P(y=3) &= 1 - \Phi(\mu_2 - \beta x_i) \end{aligned}$$

where $\Phi(\bullet)$ is the cumulative normal distribution function.

Since none of the rating agencies studied here rates the entire population of insurers, an estimated model of the determinants of insurer ratings or differences in ratings across agencies that only uses information on the sample of insurers that applied for a rating, rather than all insurers that qualify for a rating, is potentially subject to sample selection (or self-selection) bias. If there is some systematic reason why insurers without a rating have chosen not to apply for one and we estimate the rating equation based only on observations for which we have ratings, we obtain inconsis-

[25] Ordinary least squares regression assumes that the dependent variable (i.e., rating) is measured on an interval scale. In other words, it assumes that the risk differential between an AA and an A bond is the same as that between a BB and a B bond (Kaplan and Urwitz, 1979). Like Cantor and Packer (1997), we estimated our models in both least squares regression and ordered probit, and our results were qualitatively the same. We present the results based on the ordered probit model because this model is more theoretically appropriate than OLS in this situation because of the ordinal nature of the dependent variable (see Ederington, 1985; Greene, 1997).

tent estimates of the parameters and the expected error term conditional on the firm obtaining a rating is not equal to zero (Heckman, 1979; Cantor and Packer, 1997; and Greene, 1997). For this reason, we run all our models using techniques that control for sample selection bias.

In the standard sample selection model, a continuous random variable is subject to sample selection (Heckman, 1979). The standard sample selection model has been extended to ordinal random variables (Dubin and Rivers, 1990; and Greene, 1998). In the ordered probit model with selection, in addition to obtaining estimates of parameter coefficients, the correlation between the error terms from the decision to obtain a rating equation and the rating determination equation (denoted ρ) is obtained. A test for the selectivity bias is ρ=0 (Dubin and Rivers, 1990).

The ordered probit model with sample selection is:

$$d^* = \alpha z_i + u_i, \tag{2}$$
$d_i = 1$ if $d^* > 0$ and 0 otherwise,
y_i satisfies the ordered probit specification of equation (1),
$[y_i, x_i]$ is observed if and only if $d_i = 1$.

The variable d^* is a continuous unobserved variable measuring the propensity to obtain a rating, while d_i is a binary variable indicating whether a firm has a rating or not. The coefficient vector is $\boldsymbol{\alpha}$. The vector of explanatory variables is z_i. The random errors, ε_i and u_i, follow a bivariate standard normal distribution with correlation, ρ. A higher value for an explanatory variable with a positive coefficient, $\boldsymbol{\alpha}$, indicates a higher probability of obtaining a rating or multiple ratings.

Empirical Results

The Decision To Be Rated

The empirical results for the decision to be rated appear in the upper panels of Tables 4 and 5.[26] The coefficient on the inverse measure of leverage (capital to assets), is negative and significant in three of the six first stage regressions, but significantly positive in the Best equation. The three significantly negative results are consistent with the Cantor and Packer (1997) hypothesis that higher leverage is associated with greater uncertainty, which in turn is associated with a greater value of ratings. Also consistent with the uncertainty hypothesis is the positive and significant sign of the coefficient on the profitability variable in the Best regression.[27]

The results on the growth variable are surprising. If a high growth rate is associated with greater uncertainty, we would expect a positive association between growth and the probability of being rated. In fact, in five of six regressions, the coefficient on the growth variable is negative and significant at better than the .10 level.

As with some of the leverage and profitability results, the results for several other

[26] Collinearity diagnostics were performed for all models and multicollinearity does not appear to be a problem. The highest correlation coefficient is .57 (between LN(ASSET) and STATES), and the highest variance inflation factor is 2.93 (for LN(ASSET)).

[27] Our mixed results on these variables are not surprising. Cantor and Packer's (1997) results for both leverage and profitability do not support the uncertainty hypothesis.

TABLE 4
Rating Determinants Models

	Decision To Be Rated					
	Best		S&P		Moody's	
Variable[a]	coefficient	p-value	coefficient	p-value	coefficient	p-value
CAPASSET	0.6829	0.0409	–0.3414	0.3058	–2.8952	0.0000
ROA	2.9429	0.0000	1.1966	0.2395	–0.9119	0.6245
CHGNPW	–0.5108	0.0001	–0.2448	0.0915	–0.5820	0.0218
STATES	0.0149	0.0006	0.0166	0.0000	0.0219	0.0000
LOBHERF	–1.4373	0.0000	–0.9125	0.0000	–1.7836	0.0006
JUNK	–1.2973	0.6859	–1.2892	0.7123	3.9777	0.3107
STOCKS	–0.6317	0.1197	0.4233	0.1972	–0.2067	0.7082
CASH	–1.4724	0.0000	–5.1466	0.0000	–27.3857	0.0002
REINSURANCE	–0.0498	0.8164	0.5157	0.0091	0.8873	0.0054
LONGTAIL	0.1799	0.2600	0.6159	0.0022	0.9941	0.0125
PERSONAL	–0.0313	0.8401	–0.3970	0.0135	0.2130	0.4345
TRADED	0.5405	0.0010	1.0255	0.0000	1.0248	0.0000
LN(DPW)	0.1591	0.0000	0.0393	0.1688	0.1151	0.0114
	Rating Determinants					
	Best rating		S&P rating		Moody's rating	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
CAPASSET	2.6242	0.0000	1.3921	0.0284	–0.1676	0.9090
ROA	2.8748	0.0000	4.7529	0.0036	4.1311	0.2102
CHGNPW	0.2832	0.0011	0.5817	0.0036	0.2316	0.4475
STATES	0.0007	0.7063	0.0019	0.6705	–0.0032	0.6998
LOBHERF	–0.6490	0.0000	–0.2570	0.5213	–0.9190	0.4030
JUNK	–4.7768	0.0486	–15.2217	0.0032	–21.3680	0.0024
STOCKS	–1.0103	0.0000	–0.0371	0.9428	0.4401	0.6360
CASH	–1.6629	0.0000	–4.1962	0.3266	–8.3181	0.5168
REINSURANCE	0.4932	0.0000	–0.1000	0.7167	0.6727	0.2608
LONGTAIL	–0.0644	0.5718	–0.7657	0.0339	–1.0263	0.2227
LN(ASSET)	0.4023	0.0000	0.1086	0.0849	0.2632	0.0127
Rho(ε,u)	–0.5158	0.0009	–0.0248	0.9181	0.1289	0.7389

[a] CAPASSET is statutory capital divided by total assets; ROA is net income divided by total assets; CHGNPW is the percentage change in net premiums written (NPW) between 1994 and 1995; STATES is the number of states licensed; LOBHERF is the line-of-business Herfindahl index using NPW; JUNK is investments in speculative grade bonds divided by invested assets; STOCKS is common stock investments divided by invested assets; CASH is cash and short-term investments divided by invested assets; REINSURANCE is reinsurance ceded divided by the sum of direct premiums written (DPW) and reinsurance assumed; LONGTAIL is NPW in long-tail lines of insurance divided by total NPW; PERSONAL is NPW in personal lines divided by total NPW; TRADED equals one if publicly traded, zero otherwise; LN(DPW) is the natural logarithm of DPW; LN(ASSET) is the natural logarithm of total assets. Constants and threshold parameters are suppressed to conserve space.

TABLE 5
Rating Differences Models

	Decision To Be Rated					
	Best and S&P		Best and Moody's		S&P and Moody's	
Variable[a]	coefficient	*p*-value	coefficient	*p*-value	coefficient	*p*-value
CAPASSET	–0.3677	0.2729	–2.8583	0.0000	–2.5432	0.0035
ROA	1.3154	0.2041	–1.0860	0.5521	0.3827	0.8542
CHGNPW	–0.2069	0.1547	–0.4904	0.0351	–1.2860	0.0001
STATES	0.0165	0.0000	0.0217	0.0000	0.0196	0.0001
LOBHERF	–0.9452	0.0000	–1.8246	0.0002	–2.7515	0.0000
JUNK	–1.3653	0.7000	3.4456	0.3716	6.2704	0.0944
STOCKS	0.4215	0.2014	–0.1614	0.7692	–0.1853	0.7618
CASH	–5.1521	0.0000	–27.5971	0.0001	–24.6292	0.0029
REINSURANCE	0.5062	0.0117	0.8354	0.0124	0.4599	0.1884
LONGTAIL	0.5971	0.0027	1.1125	0.0050	1.0426	0.0245
PERSONAL	–0.4504	0.0089	0.2190	0.3822	–0.1481	0.6374
TRADED	1.0090	0.0000	1.0808	0.0000	0.9468	0.0000
LN(DPW)	0.0405	0.1573	0.1178	0.0099	0.1886	0.0019
	Rating Differences					
	Best vs. S&P		Best vs. Moody's		S&P vs. Moody's	
	coefficient	*p*-value	coefficient	*p*-value	coefficient	*p*-value
CAPASSET	–1.6956	0.0027	0.8796	0.5107	4.4635	0.0095
ROA	–4.2040	0.0131	0.9442	0.7701	10.6989	0.0140
CHGNPW	–0.3097	0.1734	–0.0714	0.8340	0.9780	0.0796
STATES	–0.0040	0.4046	–0.0074	0.4019	0.0045	0.6689
LOBHERF	–0.2129	0.6468	2.4699	0.0075	0.4998	0.8302
JUNK	10.4074	0.0642	9.7157	0.1300	7.2952	0.3219
STOCKS	1.5529	0.0031	0.2661	0.7821	–2.7695	0.0207
CASH	1.4659	0.7305	16.4944	0.2350	26.7480	0.1929
REINSURANCE	0.3391	0.2334	–0.3951	0.4666	–0.2328	0.7233
LONGTAIL	0.2909	0.3918	–0.4419	0.5542	2.0541	0.0884
LN(ASSET)	–0.1467	0.0357	–0.2681	0.0041	0.1994	0.0603
Rho(ε,u)	–0.1763	0.4513	–0.4137	0.2359	0.1056	0.8613

[a]See note to Table 4 for variable descriptions.

variables are also consistent with the argument that greater uncertainty leads to a greater likelihood of obtaining ratings. The variables measuring geographic dispersion and involvement in multiple lines of insurance are highly significant in all six regressions. The signs on these variables are consistent with the uncertainty argument. The investment mix variables yield mixed results. The coefficient on junk bond investment is positive and significant only once, and stock investment is never significant. However, for the investment category that clearly exhibits the lowest level of uncertainty, cash, the coefficient in all six regressions is negative and highly significant, which is consistent with expectations.

Recall that the predicted sign on the reinsurance variable was ambiguous, depending on whether the use of reinsurance was seen as reducing uncertainty by shifting risk, or increasing uncertainty by making the insurer's financial health dependent on the financial health of its reinsurers. The results support the latter argument. In four of six regressions, the coefficient on the variable measuring the percent of business ceded to reinsurers is positive and significant.

In all first stage equations except the Best equation, the coefficient on the variable measuring the percentage of business in long-tail lines is significantly positive. This is consistent both with the uncertainty hypothesis and the argument that buyers of long-tail insurance are particularly concerned about the long-term financial health of their insurers.

The measure of the percent of business in personal lines has a significant coefficient in only two of the regressions, but its sign in these regressions implies a negative association between personal insurance business and the probability of obtaining a rating. This is consistent with the arguments given earlier that commercial insurance buyers are less protected by guaranty funds and have greater demand for insurer ratings.

The variable representing whether or not the insurer is publicly-traded produces one of the strongest results. In each of the six regressions, the publicly-traded variable has a positive and highly significant coefficient. This is an interesting result, implying that insurer financial strength ratings may be important to capital markets even though insurer ratings apply to claims-paying ability rather than the quality of any particular publicly-traded security of the insurer. The results also may reflect the fact that many publicly-traded insurers already have rated debt, so the marginal cost of obtaining a financial strength rating from the agency or agencies that rate their debt is reduced.

The coefficient on the variable for the log of direct premium written is positive and significant in four of the regressions. These results are consistent with the argument made earlier that the benefit of a rating increases directly with the amount of insurance debt (i.e., policyholder obligations) being issued.

Rating Determinants

The results for the rating determinants models are presented in the lower panel of Table 4. In these models, all but one of the variables are significant in at least one of the models (number of states licensed is never significant). Only two of the variables are significant for all three rating agencies; larger insurers tend to have higher ratings, and insurers with more investment in junk bonds tend to have lower ratings. Both of these results are consistent with expectations.

Surprisingly, the variables representing size and investment in junk bonds are the only two variables that are significant in the Moody's equation, whereas the Best and S&P equations have nine and six significant variables, respectively. Thus, Moody's appears to use a much smaller number of publicly available quantitative factors in its rating process, and perhaps relies more on private or qualitative information.

Consistent with the bond rating literature, the coefficient on profitability is positive and significant in the Best and S&P models. Additionally, the coefficient on capitalization (inverse leverage) is positive and significant in these two models. Growth in premiums written is also associated with higher ratings for Best and S&P.

The coefficient on the variable measuring the percent of business in long-tail lines, which are generally considered more risky than short-tailed lines, is negative and significant only in the S&P equation. Of the remaining variables, cash holdings, stock investments, reinsurance usage, and line of business diversification are significant in the Best equation only.

Comparison of Rating Models

Looking at the results for the rating determinants models, it appears that the three models differ, in that each has a different set of statistically significant variables. However, this casual observation is not enough to determine whether the models are statistically different. Thus, we perform formal statistical tests for differences among the models. The results reported here are for the comparison between S&P and Moody's.[28] Following Ederington (1986), we estimated the two models jointly, while first restricting the μs (i.e., cutoff points) to be equal across the models but allowing the βs to vary. The results are shown in Table 6. The log likelihood for this joint model is then compared with the sum of the log likelihoods from the two models estimated separately. This difference then provides a chi-squared statistic to test for equality of the μs in the two models. With the resulting chi-squared statistic of 3.1 and three degrees of freedom, we are unable to reject the null hypothesis of equality of the μs even at the .10 level of significance. Thus, the cutoff points for S&P and Moody's do not appear to differ.

With μs constrained to be equal, the final column of Table 6 also provides chi-squared tests for differences in βs across the models. The βs differ between models at the .10 level of significance for only one variable. However, another log likelihood test was performed to test the null hypothesis that all of the βs were equal across the two models. This null hypothesis was rejected at the .01 level (chi-square = 24.8, 11 degrees of freedom), so we cannot conclude that the models for S&P and Moody's are the same.

Determinants of Rating Differences

The final set of results to be discussed is for the rating differences models. These results are shown in the lower panel of Table 5. For Best versus S&P, five variables (common stock investments, junk bond investments, size, capitalization, and profitability) are significant. Only two variables are significant in the model for Best ver-

[28] Since no evidence of selection bias was found for the S&P or Moody's models (p-values for ρ are .92 and .74, respectively; see discussion below), we do not incorporate sample selection methods into this analysis. Similar tests were performed comparing Best to S&P and Moody's. These tests decisively rejected the null hypotheses of equal μs and βs. This is not surprising, given the relatively low correlation reported earlier between Best ratings and the others (.62 and .67), compared to the correlation between S&P versus Moody's (.92).

Table 6
Rating Determinants Model With Different Coefficients and Equal Division Points

Variable	S&P's coefficient	Moody's coefficient	p-value for null hypothesis of coefficient equality
CAPASSET	1.5175	–0.3717	0.0681
ROA	4.7214	3.6263	0.7209
CHGNPW	0.5712	0.2762	0.3852
STATES	0.0013	–0.0034	0.4794
LOBHERF	–0.2567	–0.9671	0.3349
JUNK	–15.3630	–21.9693	0.3509
STOCKS	–0.1455	0.9268	0.2271
CASH	–4.1953	–5.9479	0.8811
REINSURANCE	–0.0531	0.4177	0.2949
LONGTAIL	–0.7113	–1.4468	0.3339
LN(ASSET)	0.1413	0.1817	0.2846
Division points[a]			
Mu(1)	1.5014		
Mu(2)	2.7452		
Constant	–1.1236		

[a]The constant and threshold parameters were constrained to be equal for S&P and Moody's.

sus Moody's (size and line-of-business diversification). Finally, the S&P versus Moody's model has six significant variables (common stock investments, size, capitalization, growth in premiums, profitability, and long-tail lines percentage).

Our results are consistent with the hypothesis that rating agencies differ systematically in the relative importance given to the different factors they consider. The number of significant coefficients in the rating differences models is actually rather surprising. While the bond rating literature has found numerous variables important in determining ratings, there has been little success in explaining rating differences. Cantor and Packer (1997) use measures of leverage, coverage, profitability, and size (along with industry dummies) in their models for explaining bond rating differences. In their four models, comparing Moody's and S&P to DCR and Fitch, size and profitability were each significant in only one model, while the other financial variables were never significant.

Sample Selection

One interesting aspect of our results is the relative lack of evidence of sample selection bias. In the rating determinants models, only the Best rating model yields a value of ρ that is significantly different from zero. None of the rating differences models show evidence of selection bias. Thus, the differences in ratings across agencies do not seem to be driven by self-selection issues.

CONCLUSION

This is the first study to extensively analyze insurer financial strength ratings. It examines the decision to be rated, the determinants of ratings, and the determinants of rating differences across agencies, while controlling for sample selection bias. Although corporate bond ratings have been studied extensively, the ratings analyzed here are quite distinct. Bond ratings are related to specific issues of publicly-traded debt. Insurer ratings, by contrast, relate to the overall financial strength of the insurer and are issued for mutual firms and privately-held stock firms in addition to publicly-traded stock companies.

The results demonstrate that each of the rating agencies examined uses a distinct rating model, with its own important factors and its own weights on those factors. This helps explain the large proportion of rating differences found among insurers rated by more than one agency. Particularly interesting are the differences found between Moody's and Standard and Poor's, given that Ederington (1986) did not find any differences in rating determinants for these two agencies in examining corporate bond ratings. An important implication of our results is that regulators, investors, consumers, and other interested parties should not assume equivalence of rating scales across rating agencies. Another point shown clearly is that insurer financial strength rating differences do not appear to be driven by sample selection bias, with only the Best rating determinants model providing any evidence of sample selection bias.

A number of interesting results are also found regarding the decision to obtain a rating. First, unlike Cantor and Packer (1997), we find support for the hypothesis that insurers seek ratings in order to resolve *ex ante* uncertainty about their level of insolvency risk. In addition, we find some evidence that insurers that write more business in lines with less guaranty fund protection are more likely to obtain ratings. Also, more heavily reinsured firms are found to be more likely to obtain ratings, consistent with the hypothesis that reinsurance adds another layer of complexity to evaluation of an insurer's risk. Publicly-traded firms are also found to be more likely to obtain ratings, even though insurer ratings do not relate to any specific publicly-traded security.

Our results suggest that more research dealing with the issue of insurer ratings would be valuable. For example, while this study shows that rating agencies appear to use different insurer rating models, future work could consider whether the ratings of any one of the agencies are consistently better at reflecting the true insolvency risk of the insurers it rates.

REFERENCES

Adiel, R., 1996, Reinsurance and the Management of Regulatory Ratios and Taxes in the Property-Casualty Insurance Industry, *Journal of Accounting and Economics*, 22: 207-240.

A.M. Best, 1996, *Best's Key Rating Guide, Property-Casualty Edition* (Oldwick, NJ: A.M. Best Co., Inc.).

Anthony, J. H., and K. R. Petroni, 1997, Accounting Estimation Disclosures and Firm Valuation in the Property-Casualty Insurance Industry, *Journal of Accounting, Auditing and Finance*, 12: 257-281.

Beattie, V., and S. Searle, 1992, Bond Ratings and Inter-Rater Agreement, *Journal of International Securities Markets*, summer: 167-172.

Berger, L. A., J. D. Cummins, and S. Tennyson, 1992, Reinsurance and the Liability Insurance Crisis, *Journal of Risk and Uncertainty*, 5: 253-272.

Borch, K., 1974, Capital Markets and the Supervision of Insurance Companies, *Journal of Risk and Insurance*, 41: 397-405.

Buckley, K. M., 1997, *The Relationship Between Insurance Company Debt and Claims Paying Ability Ratings* (Chicago, IL: Duff and Phelps Credit Rating Company).

Cantor, R., and F. Packer, 1997, Differences of Opinion and Selection Bias in the Credit Rating Industry, *Journal of Banking and Finance*, 21: 1395-1417.

Cantor, R., and F. Packer, 1995, The Credit Rating Industry, *Journal of Fixed Income,* 5: 10-34.

Cummins, J. D., and P. M. Danzon, 1997, Price, Financial Quality, and Capital Flows in Insurance Markets, *Journal of Financial Intermediation,* 6: 3-38.

Cummins, J. D., and R. A. Derrig, 1989, *Financial Models of Insurance Solvency* (Norwell, MA: Kluwer Academic Publishers).

Cummins, J. D., and D. W. Sommer, 1996, Capital and Risk in Property-Liability Insurance Markets, *Journal of Banking and Finance,* 20: 1069-1092.

Doherty, N., 1989, On the Capital Structure of Insurance Firms, in: J. D. Cummins and R. A. Derrig, eds., *Financial Models of Insurer Solvency* (Norwell, MA: Kluwer Academic Publishers).

Doherty, N., and S. Tinic, 1981, Reinsurance Under Conditions of Capital Market Equilibrium: A Note, *Journal of Finance,* 36: 949-953.

Dubin, J. A., and D. Rivers, 1990, Selection Bias in Linear Regression, Logit and Probit Models, in: J. Fox and J. S. Long, eds., *Modern Methods of Data Analysis* (Newbury Park, CA: Sage Publications).

Ederington, L. H., 1985, Classification Models and Bond Ratings, *Financial Review*, 20: 237-262.

Ederington, L. H., 1986, Why Split Ratings Occur, *Financial Management,* 15: 37-47.

Ederington, L. H., and J. Yawitz, 1987, The Bond Rating Process, in: E. Altman, ed., *Handbook of Financial Markets and Institutions*, 6th ed. (New York: John Wiley and Sons).

Fung, H. G., G. C. Lai, G. A. Patterson, and R. C. Witt, 1998, Underwriting Cycles in Property and Liability Insurance: An Empirical Analysis of Industry and By-Line Data, *Journal of Risk and Insurance,* 65: 539-561.

Garber, H. D., 1994, Financial Statements and Ratings, in: E. E. Graves, ed., *McGill's Life Insurance* (Bryn Mawr, PA: American College).

Greene, W. H., 1998, *LIMDEP Version 7.0 User's Manual Revised Edition* (Plainview, NY: Econometric Software, Inc).

Greene, W. H., 1997, *Econometric Analysis*, 3rd ed. (Englewood Cliffs, NJ: Prentice-Hall).

Harrington, S. E., and P. M. Danzon, 1994, Price Cutting in Liability Insurance Markets, *Journal of Business,* 67: 511-538.

Heckman, J., 1979, Sample Selection Bias as a Specification Error, *Econometrica,* 47: 153-161.

Jones, D. C., 1994, Rating Agencies Sway Insurer Investment Moves, *National Underwriter, Life, Health, and Financial Services*, 98 (19 December): 35.

Kahane, Y., C. S. Tapiero, and L. Jacques, 1986, Concepts and Trends in the Study of Insurer's Solvency, in: J. D. Cummins and R. A. Derrig, eds., *Financial Models of Insurer Solvency* (Norwell, MA.: Kluwer Academic Publishers).

Kaplan, R. S., and G. Urwitz, 1979, Statistical Models of Bond Ratings: A Methodological Inquiry, *Journal of Business,* 52: 231-261.

Klein, R. W., 1992, *Insurance Company Rating Agencies: A Description of Their Methods and Procedures* (Kansas City, MO: National Association of Insurance Commissioners).

MacMinn, R. D., and R. C. Witt, 1987, A Financial Theory of the Insurance Firm Under Uncertainty and Regulatory Constraints, *Geneva Papers on Risk and Insurance,* 12: 3-20.

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

McKelvey, R. D., and W. Zavoina, 1975, A Statistical Model for the Analysis of Ordinal Level Dependent Variables, *Journal of Mathematical Sociology,* 4: 103-120.

Millon, M. M., and A. V. Thakor, 1985, Moral Hazard and Information Sharing: A Model of Financial Information Gathering Agencies, *Journal of Finance,* 40: 1403-1422.

Moody's, 1998, *Property and Casualty Insurance and Reinsurance* (New York: Moody's Investors Service).

Moon, C. G., and J. G. Stotsky, 1993, Testing the Differences Between the Determinants of Moody's and Standard & Poor's Ratings, *Journal of Applied Econometrics,* 8: 51-69.

Munch, P., and D. E. Smallwood, 1982, Theory of Solvency Regulation in the Property and Casualty Insurance Industry, in: Gary Fromm, ed., *Studies in Public Regulation* (Cambridge, MA: MIT Press).

Pottier, S. W., 1998, Life Insurer Financial Distress, Best's Ratings and Financial Ratios, *Journal of Risk and Insurance,* 65: 275-288.

Pottier, S. W., 1997, Life Insurer Risk Characteristics and the Rating Process, *Journal of Insurance Issues,* 20: 111-130.

Schwartz, R., 1994, Good, Better, Best, *ReActions,* 14: 34-35.

Ramakrishnan, R. T., and A. V. Thakor, 1984, Information Reliability and a Theory of Financial Intermediation, *Review of Economic Studies,* 51: 415-432.

Sommer, D. W., 1996, The Impact of Firm Risk on Property-Liability Insurance Prices, *Journal of Risk and Insurance,* 63: 501-514.

Standard and Poor's, 1996, *Insurer Solvency Review, Property-Casualty Edition* (New York: Standard and Poor's).

Thompson, G. R., and P. Vaz, 1990, Dual Bond Ratings: A Test of the Certification Function of Rating Agencies, *Financial Review,* 25: 457-471.