



BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

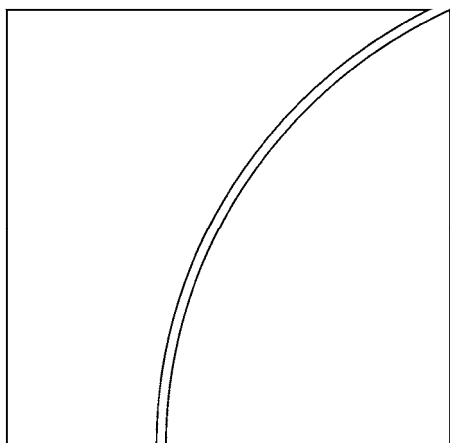
No 335

To err is human: rating agencies and the interwar foreign government debt crisis

by Marc Flandreau, Norbert Gaillard and Frank Packer

Monetary and Economic Department

December 2010



JEL classification: E44, F34, G15, N20.

Keywords: sovereign credit ratings, Great Depression, financial crisis, international bond markets

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

Copies of publications are available from:

Bank for International Settlements
Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2010. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)

ISBN 1682-7678 (online)

To err is human: rating agencies and the interwar foreign government debt crisis

Marc Flandreau, Norbert Gaillard and Frank Packer*

Abstract

During the 1930s, rating agencies took up a central role in regulatory supervision that they still have today. The proximate cause for this changeover was the economic shock of the Great Depression. Exploring the performance of rating agencies in assessing the risks of sovereign debt, an important segment of the bond market, we do not find that superior forecasting capacities can explain the agencies' growing importance.

Keywords: sovereign credit ratings, Great Depression, financial crisis, international bond markets

JEL classification: E44, F34, G15, N20.

* Graduate Institute for International and Development Studies, Geneva and CEPR, London; Sciences Po, Paris; Bank of International Settlements, Hong Kong, respectively. Part of this paper was written while Marc Flandreau was a visiting scholar with the BIS, whose hospitality is warmly acknowledged. Marc Flandreau also acknowledges the financial support Yves Mirabaud provided to the Graduate Institute for International Studies Development "Historical Imagination" project. Norbert Gaillard acknowledges financial support from the Caisse des Dépôts et Consignations. We are grateful to Magdalena Erdem for her invaluable assistance in preparing the paper. We also thank Laura Rosner for assistance with additional sources located in New York. We thank Claudio Borio, Richard Cantor, Marcello de Cecco, Jacob Gyntelberg, Jean Helwege, Michael King, Patrick McGuire, Robert McCauley, Larry Neal, Patrick O'Brien, Albrecht Ritschl, Augustín Villar, Eugene White, Mira Wilkins and seminar participants at the Bank for International Settlements, Bundesbank, London School of Economics, and the Bank of Italy conference on regulatory response to crises, for their comments on a previous draft, circulated as CEPR DP 7328. The views expressed here are those of the authors and do not reflect those of the Bank for International Settlements.

Contents

Abstract.....	v
To err is human: rating agencies and the interwar foreign government debt crisis	1
I. The interwar boom-bust cycle.....	3
II. Sovereign ratings from a business perspective	5
IV. The 1931 crisis and the era of downgrades	10
V. The performance of ratings during the interwar period: methodologies	12
VI. Rating categories and default rates.....	14
VII. Performance and accuracy ratios.....	17
a) The relative performance of agencies	17
b) Agencies vs the “market”	18
Conclusions	19
References	20
Appendix:.....	39
1. Example of Moody’s ratings (Dominican Republic, 1927, p 444 of the Manual):	39
2. Formula for the accuracy ratio (AR).	39
3. Formula for the yield-implied ratings	40

To err is human: rating agencies and the interwar foreign government debt crisis

On 14 September 1929, an ad placed in *The Saturday Evening Post* by *Standard Statistics*, a US rating agency, read: “Today You Need not Guess [...] History sometimes repeats itself – but not invariably. In 1719 there was practically no way of finding out the facts about the Mississippi venture. How different the position of the investor in 1929!” The ad went on to explain that modern investors were fortunate to be able to rely on the opinion of rating agencies such as Standard Statistics. The claim may sound hollow given the financial collapse that was to come and ironic in the light of more recent discussion. Ever since the East Asian crisis of 1997 credit rating agencies have been a lightning rod for criticism when financial markets were perceived to fail. In the midst of the financial collapse of Thailand, Indonesia and South Korea, many observers blamed rating agencies: after having been unable to predict problems, they were perceived to have amplified business fluctuations through hasty downgrades. The modern conventional wisdom among many scholars today is that rating agencies display a tendency to be reactive to financial crises rather than anticipate them (Reisen and Von Maltzan (1999), Ferri, Liu and Stiglitz (1999), Reinhart (2002), Partnoy (2006) and Benmelech and Dlugosz (2009)).

Relatively little is known about the prehistory of ratings, and the classic account of the development of the securities rating business is still Harold (1938). However, the topic has attracted considerable research interest more recently. Economic sociologists and historians have made important inroads into the study of business rating by so-called mercantile agencies, as the 19th century predecessors of rating firms were known (see Carruthers and Cohen (2006) and Lauer (2008)), as well as of business ratings in other countries (Berghoff (2008)).¹ References also include Sylla’s excellent overview of the history of security certification in the US (Sylla (2002)), Flandreau’s paper on the origins of sovereign debt rating in France in the 19th century (Flandreau (2003)), and Partnoy’s work on the perverse effects of regulation on ratings and rating quality (Partnoy (2006, 2009)).

This paper makes a contribution to this emerging literature on the history of ratings by focusing on the assessment of foreign government debt by US rating agencies during the interwar period. There are two reasons for this focus. First, it was during the interwar period that ratings became involved in regulation. This happened as a direct response to the debt crisis: when all bond prices plummeted in the midst of the financial turmoil of the summer of 1931, the Office of the Comptroller of the Currency (OCC) – a department of the United States Treasury in charge of supervising nationally chartered banks – made an emergency decision to use credit ratings-based formulae to book the value of US national banks’ bond portfolios.² From that point on, ratings became the main instruments through which regulators supervised banks’ risk exposures – which suggests that they were seen then as part of the solution, not as part of the problem. Second, looking at the government debt crisis may add a useful perspective on what remains one of the most (perhaps the most) violent foreign debt disaster in financial history. Between 1931 and 1939, more than half of sovereign borrowers who had issued in New York during 1920–29 defaulted. The episode, an integral part of the catastrophic interwar financial system dislocation, has not yet been studied from this vantage point, although many other aspects of the crisis have been

¹ We also mention here the contribution of business historians such as Norris (1978).

² See Harold (1938), p 27 for a discussion.

discussed in detail.³ In previous narratives, rating agencies are either ignored or are viewed as impartial judges of the debacle.

At the broadest level, this article belongs to a tradition of papers that has discussed whether the public could or did foresee interwar financial problems (Dominguez, Fair and Shapiro (1988), Klug, Landon-Lane and White (2005) and Favero (2007)).⁴ The sovereign debt market is a particularly relevant venue to ask this question. It was reported to be where abuse of investors by underwriting banks was most rampant (US Senate 1932). Afterwards, one of the most vociferous critics of sovereign debt origination (which was perceived to have malfunctioned during the 1920s) was Max Winkler, head of Moody's sovereign ratings unit until 1928.⁵ In 1933, Winkler claimed that, as early as in July 1925, he had emphasised in a special report released by Moody's Investors Service that banks were "unethical" and that the "many circulars, descriptive of foreign loans issued by prominent Banking and Financial Firms, contain[ing] statements which are either incorrect or decidedly misleading", would later cause problems.⁶ However, the precise wording he used at the time suggests that even he did not anticipate the extent of the catastrophe, for he wrote that "the *rather small number of defaults* on foreign issues [to be expected] *would be even smaller* if the methods employed by the bankers were less unethical".⁷

Beyond the broader question of investor foresight also lie a number of additional interesting issues. One that has received much attention lately is the impact of the dynamics of competition on ratings and rating quality. The theoretical argument is that competition among providers of ratings can lead to a general deterioration of the quality of ratings known as "rating inflation".⁸ One can think of the intuition behind this prediction as being a variant of Gresham's law: Because issuers only need to show they are good by one standard, and because they pay for that, poor-quality ratings drive out the good, resulting in rating inflation (similar to currency depreciation and inflation as described by Gresham's law). An important take-away from this literature is that the effects of competition can be distorted by regulation. The regulatory licence created by the use of ratings in financial supervision can then be seen as a source of financial instability.⁹ Proposals to reform the use of rating agencies' output are thus motivated by the view that, if rating agencies had to earn their reputation, they would be more careful and resist rating inflation. On this account the record of the interwar period is of peculiar value, since this period was before (at least until the crisis in 1931) the emergence of modern day ratings-based financial supervision. We think it is interesting to study the

³ See Winkler (1933), Lewis (1938), Madden et al (1937), Rippy (1950), Mintz (1951) and later Wigmore (1985), Morton (1939), Eichengreen and Portes (1986), Eichengreen and Sachs (1986) and Eichengreen (1989).

⁴ The anecdotal motivation for this literature is the well advertised forecast made on September 5, 1929, by Roger Babson that "wise investors will pay up their loans and avoid margin speculation at this time because a 'crash' of the stock market is inevitable", *The New York Times*, 6 September 1929. Babson ran a forecasting boutique and had also personal investments in the rating industry with Moody's and Poor's. He is said to have later recapitalised Poor's after it went bankrupt in 1931. His forecasts, which were blamed by some to have caused the crash, were in contrast with Irving Fisher's remark on the "permanent plateau" in stock prices.

⁵ Winkler (1933); Winkler was a member of Moody's editorial board between 1922 (the year of the creation of a formal rating department at Moody's, managed by W Barrett Brown) and 1927. He resigned from Moody's in 1928 to set up his own consultancy.

⁶ Winkler (1933), p 58. For similar statements, see pp 58–64.

⁷ Winkler, 1933, p. 63. Our italics.

⁸ See Skreta and Veldkamp (2008), Benmelech and Dlugosz (2009) or Bolton, Freixas and Shapiro (2009) for recent theoretical contributions.

⁹ For an elaboration of the risks of investors relying exclusively on credit ratings in the context of structured securities, see Committee on the Global Financial System (2005, 2008).

agencies when they had to rely on reputation alone for their business – and at a time when, among several other institutional differences, they were charging investors and not issuers.

In this article, therefore, we describe the record of US rating agencies before and during the interwar sovereign debt crisis. To this end, we have gathered and analysed data on ratings and financial markets during the sovereign debt boom-bust cycle of the 1920s and 1930s. Focusing on foreign government securities listed on the New York Stock Exchange, we compare ratings' reliability among different agencies and between ratings and market yields. We document a large degree of procyclicality of ratings over the period. Perhaps more surprisingly, rating agencies do not appear to have performed particularly well relative to financial markets in forecasting the approaching mess: when we compare the predictive power of agency ratings with that of synthetic ratings based on market yields, we find little that suggests strongly superior performance. Our results leave open the reasons for the emergence of ratings as regulators' preferred instrument, for superior performance does not appear to have motivated the initial regulatory use of ratings.¹⁰

The remainder of the paper is organised in seven sections that belong to two broad parts. The first part comprises three sections that provide background information. Section I documents the interwar foreign debt boom and bust, section II surveys the development of the rating business during that period, and section III explores the rating output focusing on the 1920s. The second part comprises four sections and tracks the agencies' behaviour during the interwar debt crisis. Section IV shows that 1931 was a turning point, and sections V to VII rely on a variety of techniques to assess the agencies' performance. We end with conclusions.

I. The interwar boom-bust cycle

Between 1920 and the early 1930s, the New York bond market experienced a classic boom-bust debt cycle. Total bond issuance grew rapidly through much of the 1920s. Figure 1 shows the evolution for both domestic and foreign sovereign and sub-sovereign bonds. Total foreign issues peaked in 1924, when they were more than 25% of domestic corporate issuance. Domestic corporate issuance reached a maximum in 1927. Bond issuance contracted dramatically after 1930, following the stock market crash and credit crunch.

Foreign government bonds were a new line of business in New York. Prior to 1915, the US experience in sovereign lending and underwriting was limited.¹¹ Following the outbreak of World War I, the United States began playing a leading role through the financing of Great Britain and France, beginning with the Anglo-French war loan of 1915. Several short- and long-term bonds were floated during 1915–17 on the NYSE.¹² The size and scope of New York issues increased after the war, and came to include sovereign, sub-sovereign, government-guaranteed and corporate securities from Latin America, Europe and the Far East (see Figures 2.a and b). One of the regions most benefiting from US government support of banks' foreign activities was Latin America.

To explain this expansion, some authors have emphasised the shifting position of the US current account from a structural deficit to a structural surplus and the accommodative

¹⁰ One caveat is that the evidence we have is only for sovereign bonds, and complete discussion of the matter would involve the daunting task of collecting and exploiting data for corporate ratings as well.

¹¹ As of 31 December 1914, the total amount of outstanding foreign securities remained modest compared to figures for leading European markets. See Lewis (1938), pp 332–38 and 338–50.

¹² See *Moody's Manual* (1918) for details.

monetary policy that was associated with the Gold Exchange Standard. Another explanation is the success of the Liberty Bonds issued between April 1917 and September 1918.¹³ Another popular claim is that US policymakers may have helped by liberalising issuance procedures in 1923.¹⁴ Some authors have pointed to the deterioration of credit standards of some or most underwriting banks.¹⁵

An often-discussed issue is the extent to which investors were able to come to terms with the difficult notion of debt sustainability. The key issue here was whether countries “over-borrowed” or not. For instance, in a 1927 memorandum, S Parker Gilbert, US Agent General for Reparation Payments in Germany, warned about “the accumulating evidences of overspending and over-borrowing on the part of the German public authorities”.¹⁶ Similarly, US Department of Commerce Statistician Corliss told the US Senate Committee on Finance hearings on the sale of foreign government bonds in 1931–32 that he would have wished to provide statistical evidence to inform any discussion of over-lending.¹⁷ Authors have emphasised the role played by ignorance (investors had little knowledge of the history of sovereign debt as there had been almost no sovereign bond defaults in the 1920s and few before), myopic behaviour (the great profitability of past foreign bond issues was viewed as sustainable), and miscalculations in sustaining the boom.¹⁸

Foreign issues on the NYSE peaked in 1927, and issues stalled following the Crash of 1929, as investors began questioning the creditworthiness of sovereign and sub-sovereign borrowers. In the wake of the Central European financial crisis, foreign bond prices plummeted. The next stage was a wave of defaults, which began in 1931 with Bolivia, Brazil, Chile, Colombia, the Dominican Republic, and Peru, followed in 1932 by Argentina, Austria, Bulgaria, Costa Rica, Greece, Guatemala, Hungary, Panama, Salvador, Uruguay, and Yugoslavia. In 1933 most German sovereign and sub-sovereign bonds went into default.¹⁹ By September 1939, 39% of the sovereign bonds issued on the NYSE in the 1920s had defaulted, accounting for 55% of the issuance volume over that period. And more than four-fifths of the number of sovereign bonds issued in the 1920s that defaulted in the 1930s did so between 1931 and 1933.

A substantial literature has discussed reasons for the bust. Conventional interpretations try to distinguish between unwillingness and inability to pay. Citing inability to pay, researchers note unsustainable debts owing to international conditions.²⁰ In particular, many emphasize the contribution of sharp declines in commodity prices after 1929, which undermined the fiscal base of sovereign debtors and caused current account problems. Such problems were amplified by currency mismatch, since debt service was in dollars when revenues were in

¹³ These debt swaps between creditworthy US authorities and financially weak allied countries would have encouraged investors to become familiar with foreign loans while at the same time giving them a false sense of security. On these matters see Young (1930), Lewis (1938), Madden et al (1937) and Eichengreen (1989).

¹⁴ See Edwards (1928) for an early discussion.

¹⁵ Madden et al (1938) and Mintz (1951), US Congress, 1932, pp 1268–72, and Rippy (1950), pp 240–1.

¹⁶ Mintz (1951), p 77.

¹⁷ US Senate (1932).

¹⁸ Mintz (1951), pp 70–1. Only two Brazilian bonds issued in the 1920s defaulted during the same decade (see *Moody's Manuals*). See Morrow (1927) for the attraction of foreign government bonds on American individual investors. Winkler (1933), pp 54–9, Rippy (1950), p 238, Mintz (1951), pp 70–1 73–8.

¹⁹ See *Moody's Manual* (1935).

²⁰ Lewis (1938), pp 382–7. See Winkler (1933), pp 80–1. Contreras (1990) emphasises the large share of unproductive, military spending in Bolivia.

domestic currency.²¹ We intend to add to this narrative a study of the record of rating agencies.

II. Sovereign ratings from a business perspective

When our story begins in the late 1910s, securities rating was a thriving branch of financial journalism. Corporate rating already had a long history, dating back to 1857 when Bradstreet's *Book of Commercial Reports* was first released.²² The expansion of corporate rating by so-called mercantile agencies was accompanied by a series of trials from which the predecessors of modern rating agencies emerged reinforced thanks to an interpretation of the principle of "qualified privilege", which enabled these firms to produce and sell advice in the form of grades that summarised creditworthiness as raters perceived it.²³

At about the same time mercantile agencies became ingrained in the US corporate credit system, systematic descriptions of securities gained popularity. A variety of volumes documenting the characteristics of bonds emerged, although no ratings were yet given. These included Henry Varnum Poor's *Manual of the Railroads of the United States* (first published in 1868),²⁴ the *Manual of Statistics* (first published in 1879), and Moody's *Manual of Industrial and Miscellaneous Securities* (first published in 1900). These fairly thick volumes (the *Manual of Statistics* was "1040 pages but handy in both bulk and arrangements", said the ads²⁵) were used by banks, insurance companies, industrial corporations' treasurers and probably by some individual investors who consulted them at their bank.

Securities rating as we know it was started in 1909 when John Moody (after having gone bankrupt from other investments in the 1907 crisis) was forced to sell his earlier *Manual*.²⁶ He then reinvented the *Manual* with the *Analyses of Railroad Investments* and spiced up the product by adding ratings of railroad securities. The idea of providing ratings for securities (in effect of government securities) had been toyed with in the past by Nash (1889) and by French bank Crédit Lyonnais (Flandreau (2003)), but never on the scale adopted by Moody. Poor's followed suit in 1916. Fitch and Standard Statistics then both began providing ratings in 1922. This strongly suggests that investors valued summary assessments of outlooks.

In the interwar period, competition for supplying ratings to the public became active, and the firms involved were large and profitable.²⁷ A general characterisation of the competitive landscape could be as follows: Moody's was the incumbent; Fitch and Poor's were followers. Standard Statistics, which started as a "card index" provider in 1900 (thus tracking information in firms for its customers) expanded aggressively and became second to

²¹ Eichengreen and Portes (1986), p 612, James (2001).

²² Bradstreet's Books did contain a list of companies; there was a complementary system of keys to reports so that the listings could be converted to ratings. See Norris (1978).

²³ See Sandage (2005). "Qualified privilege" provided an exception to libel law for providers of socially useful information.

²⁴ This annual provided details on the performance of railway lines and companies; Chandler (1956).

²⁵ *New York Times*, 2 July 1904.

²⁶ *Moody's Manuals* (1933).

²⁷ Moody's went public in 1928. Standard Statistics advertised having more than 1,000 employees.

Moody's in terms of the numbers of ratings issued (it would later merge with Poor's, which experienced financial difficulties during the 1930s).²⁸

Thus, while it is not quite the case that ratings were an "American financial innovation" (Sylla (2002)), it is true that the innovation took hold in the United States. The provision of ratings can be thought of as one aspect of the increased commercialisation of investment advice, which the newly created statistical agencies were providing. After 1900, providers of financial information increasingly became "advisory agencies" and included investment recommendations.²⁹ As a result, the key dimensions in which rating firms competed were coverage and timeliness.

Table 1 provides a summary of what the agencies did in the interwar period, when, and for what price, with a particular focus on 1930.³⁰ Rating agencies covered securities through a variety of tools. On the one hand there was the reference instrument, which in the case of Fitch, Moody's and Poor's was the manual. For Standard Statistics, the reference tool was a card system, although some manual-like instruments were later produced, apparently as a way to attract customers to Standard Statistics. Comparison of the number of pages in the volumes for the three agencies with manuals reveals Moody's to have had the most. Its thicker books were also more expensive. Moody's had also distinct manuals for Municipal & Government, Banks and Finance Companies, Industrials, Public Utilities and Railroads while competitors had more aggregated products.

Coverage by publications of Moody's, Poor's and Standard Statistics was close to exhaustive (missing securities would usually be covered in the next volume), though Moody's tended to be the most comprehensive. For instance, the number of sovereign securities listed on the NYSE at the peak of the interwar foreign debt boom in 1929 was 116: Moody's covered 115 of them, Standard Statistics 114, Poor's 113, and Fitch only about 105 (or about 90%).

Frequency of updates was an instrument for marginal differentiation. To this end, Moody's developed its *Investment letter*, a formula that was later adopted by Fitch and Poor's. Standard Statistics then started its daily updates in 1922.³¹ From our perusal of sources and advertisements, we observe a gradual increase in the frequency of updates over time: weekly, then semi-weekly, then daily products were launched. Prices for these services were similar across agencies during the interwar period.³²

Competition also led agencies to offer new financial products to meet the increasing scope of investors' interests. For instance, Moody's started reviewing sovereign and sub-sovereign entities in the annual *Governments and Municipals* handbook in 1918, at a time when the New York market for such securities was still in its infancy. The move was followed by other

²⁸ A "card index" was a system that enabled subscribers to monitor securities through cards, which were received by post and could be filed in corresponding drawers. The other important firm in this business was Babson Systems.

²⁹ For instance, in 1907 the Stock Department of Babson System published *Investments: What and When to Buy. The use of statistics in accumulating a fortune* (published as Lamar (1907)). Other outputs that emerged at the same time were business forecasts, which were to be eagerly covered in the press. See Babson (1910) for a survey of financial information.

³⁰ Constructing this Table has involved a fair amount of conjecture and we have tried to make this as explicit as possible in the notes and references.

³¹ This service was slightly more expensive but it may have satisfied an important need, for it eventually came to be adopted by all other agencies.

³² In 1929, the Fitch subscription stood at USD 200, compared to USD 175–275 for Moody's, depending on the subscriber's willingness to receive weekly information and ratings on one sector only (eg foreign government bonds) or all sectors covered. Fitch, for instance, in 1933 launched a daily publication for USD 180 per year, similar in cost to the daily coverage by Standard Statistics.

rating agencies: Poor's began sovereign ratings in 1922, Fitch and Standard Statistics in 1924. Other markets of greater or lesser importance to US investors could also be covered by rating agencies. One prominent example is the London market, where many non-dollar international sovereign bonds were traded which Moody's covered, though not the other agencies. Moody's also rated more domestic currency local bonds not listed in either New York or London than the other agencies.³³

Firms also attempted to differentiate through prestige. Standard Statistics posted names of professors from famous universities as part of their board of "contributing editors".³⁴ Moody's did employ Max Winkler and boasted his PhD in economics. While Standard Statistics was an established competitor in the business of selling statistical information, it was new to the game of selling ratings. Hiring academic economists may have been meant to compensate for its lack of reputation or perhaps to position it as a provider of cutting edge "scientific" research.³⁵ Standard Statistics also advertised a large network of branches signalling a strong global presence, while Moody's established its international reach very early on with a branch in London. Fitch's international expansion came later.

III. Rating foreign government debt in the 1920s

This section first surveys ratings symbology, and then moves on to discuss the agencies' rating of foreign government debt during the 1920s. Collecting historical ratings today is not an easy task. In general, for want of convenient access to the agencies' higher-frequency publications, which we found to lack comprehensiveness or be difficult to use, we settled for annual data collected from the more widely available manuals (see Table 2 for details).³⁶ Comparing ratings across agencies raises methodological issues. Unlike today's readily comparable scales, rating symbols were not supposed to match one another, and the firms insisted on this. The A, B, C and [D] ranking was common to all four agencies, but different firms used different keys for granularity. For instance, to decompose the top A category, Fitch used AAA, AA and A; Moody's Aaa, Aa and A; Poor's A**, A* and A; and Standard Statistics A1+, A1 and A. Complications included Moody's lack of a letter D and Poor's initial use of a "super-rating" category above A** that was "practically inapplicable" to most securities and discontinued in 1939 (Harold (1938), p 74).

Harold (1938) gives product differentiation as one likely reason why agencies did not initially seek consistency. Differentiation could enable agencies to interpret own forecast errors as owing to the different "meanings" of their symbols.³⁷ That said, the use of ratings in regulation in the 1930s implied at some level that they could be compared, and indeed, there appears to have been a consensus in regulatory circles on this. The 1931 ruling by the Office of the Comptroller of the Currency made an official reference to the top four rating categories

³³ Grieser (1938) describes how Moody's made an alliance with *The Economist* during the interwar period to cover products for England-based investors (without ratings, however, which were not part of the business standards in Great Britain), and this may have helped it provide coverage of non-US instruments in its manuals.

³⁴ In February 1922, the Standard Daily Trade Service publication indicated "Contributing editors: Eugene E. Agger, Associate Professor of Economics, Columbia University; Lewis H. Haney, Director, Bureau of Business Research, NYU; DR Scott, Professor of Economics in the University of Missouri; Consulting economist: H.J. Davenport, Professor of Economics, Cornell University".

³⁵ Standard Statistics (1931).

³⁶ For Standard Statistics, annual data were worked out from monthly publications containing updates; we recorded as rating outstanding for a given year the latest rating mentioned in the series of monthly publications and assigned it the date of December of that year (as it was the current rating at that time). During a later visit to the Baker library, it proved impossible to get access to more granular data, reflecting the challenges associated with this kind of research.

³⁷ The role of signal precision in reputation acquisition is studied in Bolton et al (2008).

as being the safest ones, which thus implied there was a correspondence across the agencies.³⁸ This usage may then have provided incentives for agencies to make their ratings categories more consistent with each other.³⁹ In any case, Harold (1938) draws on the language used by agencies to establish what he calls a “majority interpretation” of the correspondence across the individual agency ratings: at the top, for instance, Fitch described AAA securities as “highest”, Moody’s Aaa were “highest”, Poor’s A** “very high” and Standard Statistics’ A1+ “highest class”. At the bottom, the D letter meant “slight or nil” to Fitch, “practically valueless” to Poor’s and “doubtful value” to Standard.⁴⁰ We reproduce the rating scales in Table 3 and have accordingly added one column showing correspondence with modern keys (which have a finer granularity). Harold’s majority interpretation is the criterion we used for aligning the scales.⁴¹

Having collected ratings for each agency, we are able to track the evolution of the grades they gave. Figure 3 begins with the record of the 1920s. It shows the breakdown of new foreign government securities brought to the NYSE during the 1920s. Results for this figure are shown using Moody’s, as it was the agency with the longest and most exhaustive coverage and was thus “most representative” (the other agencies had a similar behaviour). The figure shows a deteriorating quality of the portfolio of newly issued sovereign bonds in New York before 1929. The proportion of new securities rated Baa and below rises steadily across time to more than 70% in 1929. To be sure, securities rated Baa or their equivalent were still described as “good”, and later on this category belonged to the top four notches that came to be known, owing to the privileges it received from regulators in 1931 as the “investment grade” category. On the other hand, investment handbooks and the agencies themselves emphasised that below the A-rated group of securities, risks were significantly higher (motivating the change of letter). This is an important result as it suggests that rating agencies saw the decline in the quality of the bonds they were rating but did not anticipate that these risks would morph into a disaster of such a global scale.⁴²

A useful way to get a sense of the economic indicators that were relevant to rating agencies when they gave grades is to take a closer look at the ratings of Moody’s, which has left more abundant material. Just like today, rating agencies were rather cautious in sharing information on how grades were given, but Moody’s Investment Letter (1919) and Moody’s Investors Service (1924) provide details on the variables it deemed important. In 1919, when foreign government rating was still a very new business, the list reflected heavily post-war concerns: it included “national income”, “population changes caused by the war” and the evolution of the “earning power of peoples”. By 1924, the list had not only become more exhaustive, but had also acquired a definite “modern” ring. There were 25 criteria that could

³⁸ Note that a mysterious feature of the OCC ruling is that it referred to Moody’s, Standard Statistics, Fitch and one “Bond and Quotation Service” apparently having a rating system based on A-3, A-2, A, B-3, etc. See Office of the Comptroller of the Currency, 29 August 1931. We have been unable to locate this rating agency, which Harold (1938) never mentions. We came across a second reference to it in one Fed report (Federal Reserve Committee on Branch, Group, and Chain Banking (1932)), which uses ratings “assigned by important rating services, such as Moody’s, Fitch, Standard Statistics, or Bond and Quotation Service”.

³⁹ We do not know whether the fact that Poor’s was not included in the OCC’s decision of 1931 had to do with the existence of its “super-ratings”. This may explain why Poor’s deleted its super-ratings (A****, A****, and A***) in the late 1930s. The reduction of Poor’s rating range downgraded top bonds (US government, British, and Canadian bonds in 1938 and 1939) and amalgamated them into the new current highest rating.

⁴⁰ There were also on occasion discrepancies across firms for intermediary ratings. For instance, Fitch’s B was “speculative” while Standard Statistics’ B was “semi-speculative” (“speculative” was C1+).

⁴¹ In most of the analysis that follows, a strict correspondence across ratings is not needed, since several of the comparisons we make are about ordinal properties of ratings, not absolute correspondence.

⁴² This is consistent with earlier claims in the literature suggesting that the low-quality bonds came later in the credit cycle (Mintz (1951)).

be grouped into seven broad categories. A first group of indicators captured the country's degree of industrialisation. A second reflected institutional, political, and geopolitical stability. A third was for the track record (debt repayment in the past etc.); A fourth reflected monetary stability; a fifth foreign trade; a sixth fiscal balance; and a seventh stood for the debt burden (relative to wealth or revenue).⁴³ These categories also parallel to a considerable extent research for the modern period (Cantor and Packer (1996)), which found that most of the variation in the average of Moody's and Standard & Poor's sovereign ratings as of September 1995 could be explained by five variables: per capita income, inflation, external debt, level of economic development and default history.⁴⁴

We seek to identify the main determinants of the grades given during the 1920s, ie the factors in the minds of Moody's analysts (and, by extension, the analysts of other agencies) when they delivered the ratings.⁴⁵ Poring through the data in *Moody's Manuals*, we extracted a number of indicators reflecting the quality of foreign government debt that enabled us to replicate the exercise in Cantor and Packer (1996). We dropped some variables that the Moody's 1920s avatar defined more loosely or which are harder to measure, and focused on those variables for which there is a clear statistical definition.⁴⁶ These variables are used to estimate what weights were given to each indicator when producing ratings for outstanding debts at two benchmark dates, viz. 1925 and 1929 (the middle and the peak of the foreign debt boom).

The explained variable is the rating a country received at the benchmark date (1925 or 1929).⁴⁷ The explanatory variables are the variables in Cantor and Packer (1996) constructed for the 1920s sample: per capita income (measured as the wealth per capita in 1921 and 1925); growth rates (measured as the average annual GDP growth on a year-to-year basis for 1921–23 and 1925–27, in per cent); inflation (the average annual consumer price rate for 1921–23 and 1925–27, in per cent); fiscal balance (measured as the average annual central government budget surplus relative to gross wealth for 1921–23 and 1925–27, in per cent); external balance (measured as the average annual trade surplus relative to gross wealth for 1921–23 and 1925–27, in per cent); external debt (measured as the foreign currency debt relative to exports in 1923 and 1927, in per cent); a development indicator (Moody's classification as a "manufacturing country" in 1925 and 1929); and default history (the default on foreign currency debt since 1900 and since 1904, respectively).⁴⁸ For ease of interpreting the coefficients, ratings were quantified in such a way such that high values of

⁴³ See Gaillard (2008) for details. An eighth criterion was for the "legality and validity of issue". Moody's never claimed to have detected a bond whose issue was "illegal", although it may have used this criterion to vent its spleen about "unethical bankers".

⁴⁴ See Jüttner and McCarthy (2000), Bhatia (2002), Afonso (2003), Sy (2004) and Moodys (2006) for later discussion.

⁴⁵ This enables us to go beyond some studies such as Tomz (2007, p 74) who seeks to trace the type of orthodoxy at work in foreign bond markets and embedded in the wisdom of "investment experts" such as "bond rating agencies including Moody's and Fitch" by providing a qualitative discussion of the variables that mattered for these. For a related investigation looking at the internal consistency of historical rating systems, see Flandreau (2003).

⁴⁶ For instance, it appears that Moody's had strong racial prejudices ("racial characteristics" were routinely associated with wealth) and it would be possible, in principle, to test whether such views had an impact on the grades its delivered. An alternative approach would be to infer from Moody's own manuals a list of variables and test them using the same regression framework. See Flandreau (2003) for an illustration of this methodology.

⁴⁷ To account for the fact that rating in the interwar period focused on individual securities rather than countries, we have taken the average grade each country received for its portfolio of securities. Generally, grades for different securities by the same country were identical or similar, and the transformation is harmless.

⁴⁸ The logarithms have been taken for the per capita income and inflation variables.

so-called good variables such as wealth should be associated with higher ratings, and thus yield positive coefficients in the regression, and high values of bad variables such as debt burdens should be associated with lower ratings, and thus yield negative coefficients.

We use two regression techniques. The first rests on the conventional transformation of grades into numbers enabling implementation of straight OLS regressions (see Cantor and Packer (1996)). We also consider alternative estimating procedures relying on the so-called ordered probit technique. Ordered probit is a latent variable technique enabling researchers to model rank-ordered data. It relies on maximum likelihood techniques. Results are presented in Table 4: the first column gives results from Cantor and Packer (1996), while the other columns report results for 1925 and 1929 using the two techniques mentioned above.

The results yield coefficients on the variables that generally have the expected signs. Income per head and a good track record systematically are associated with higher ratings, while inflation and external debt are associated with lower grades. Interestingly, inflation and the track record are no longer significant in the second cross section (1929), which could be interpreted as suggesting that rating agencies may have become less worried about these variables. The growth and development indicators seem to have had negligible effects throughout (perhaps owing to multicollinearity). Lastly, we find that fiscal and external balance either were non-significant or (especially in the 1929 cross section) had the wrong sign. This suggests that the rating agencies did not place trade or fiscal imbalances at the top of their list of risk factors; they may have thought of these variables as endogenous, eg only “safer” countries were able to build up large trade or fiscal deficits. Finally, the table is strongly suggestive of the stability of the determinants of ratings over the very long run. Of the five variables that are significant in the modern period (income per head, inflation, external debt, the development indicator and default history), four are significant with the correct sign in most of the historical regressions.⁴⁹

Two main conclusions emerge from this section. First, one did not need to wait for the work of later economists such as Mintz (1951) to identify the deteriorating quality of foreign governments’ securities issued on the NYSE during the 1920s. Rating agencies did report on this, although the reports they gave were hardly a reason to panic as most grades remained above the investment grade line (Baa/BBB). Second, we found that, broadly speaking, the determinants of ratings during the 1920s were not that different from those that have been more recently estimated.⁵⁰

IV. The 1931 crisis and the era of downgrades

When the debt crisis struck in 1931, rating agencies, in line with the general collapse of market sentiment, revised their grades massively downwards. This is captured in Figure 4, which shows the evolution of ratings for the population of outstanding securities for two agencies (Moody’s and Fitch).⁵¹ Since new issues of foreign government securities were essentially discontinued after 1930, the sharp deterioration in rating composition reflected a

⁴⁹ Two of the variables, external debt and default history, were also key determinants of pre-WWI ratings (see Flandreau (2003)).

⁵⁰ The bigger intercept parameter for the modern period (though not significantly larger) suggests that earlier ratings were not systematically more optimistic.

⁵¹ As the lists of securities rated by the two agencies were not identical, differences may represent differences in composition of issues rating as well as differences in the ratings assigned to common borrowers. However, the number of securities rated by some agencies and not others is relatively small, and the charts do not differ greatly if only jointly rated issues are used.

net increase in downgrades of outstanding bonds. For Moody's, while only about one-fifth of the issues were rated Ba and below (ie "speculative grade") in its January 1930 manual, that fraction had risen to 50% by January 1933. The case of Fitch shows an even more dramatic change: while on average less than 10% of their August 1930 ratings were BB and below, by August 1933 more than 60% were in that category.

Working with Moody's higher frequency publications (the weekly *Investment letter*), we were able to track the precise date when Moody's own downgrades accelerated. After a summer in which the agency started downgrading some securities as sovereign defaults were beginning to emerge, massive downgrades across the board were announced by Moody's in just one week. This was on 28 September 1931. On that date, Moody's downgraded 46 of the 81 securities for which we have complete information – ie 57% of outstanding securities. The downgrades were themselves occurring in a market in complete retreat. Constructing an event window around 28 September 1931, and using bond quotations from the *Wall Street Journal*, we find that between 25 September and 2 October, the yields on bonds Moody's downgraded increased on average by 102 bps. Securities that were not downgraded held their ground (an average decrease of 22 bps, suggesting a flight to quality). The large number of events that occurred during the period makes it impossible to identify causation – let alone to measure the effect Moody's downgrades might have had on the market. But we note that Moody's downgrades occurred in the week that followed the sterling crisis (Wigmore (1985)). In any case, the speed with which the change of mood occurred is mesmerising: while ratings had held up after the 1929 crash and then deteriorated slowly during the first eight months of 1931, subsequent revisions came with a bang in an atmosphere of contagious panic.

Moody's change of mind was not isolated – but since it is the only agency for which we can provide high-frequency analysis, comparisons must be made carefully. The annual data we have do show that various agencies reacted in somewhat different ways within a common broad pattern. Figure 5 tracks upgrades and downgrades for the four agencies. Because of the different publication dates of the various agencies' volumes (see Table 2), downgrades associated with the September 1931 crisis would show up as 1932 for Fitch and Poor's and as 1931 for Standard Statistics.⁵² The chart shows that between January 1931 and January 1932, Moody's downgraded 80% of the outstanding securities, compared to the 57% it downgraded in just one fateful week. Fitch's manual, which was released in August, had the highest number of pre-crisis downgrades. On the other hand, Fitch's downgrades were to be even higher (90%) during the later period, which covered the September 1931 crisis. Poor's (the ratings of which came out in March) had also already begun to revise its ratings in advance of the crisis (see 1931 in Figure 5, ie March 1930 to March 1931), but then again, the peak came between March 1931 and March 1932, when about 50% of securities were downgraded. Standard Statistics seems to have been the least reactive to the crisis – although it would be interesting to know with greater detail its behaviour during September 1931: it downgraded 30% of the securities between December 1930 and December 1931, and 50% again between December 1931 and December 1932.

Of course, the need for these downgrades was in part a function of pre-crisis ratings. For instance, we see that Fitch was the agency that made the largest number of *upgrades* during the 1920s: its subsequent revisions represented reversals of its earlier opinions. Standard Statistics may have been more confident in the grades it had given earlier, and thus held its ground somewhat more consistently. However, we see that the foreign debt crisis eventually shifted the views of all the agencies, and the number of downgrades increased from 1931 to

⁵² Moody's came out in January, Fitch in August, Poor's in March, and we have collected Standard for December.

1934. Clearly, with the crisis of 1931, the rating agencies awoke to a new, more frightening world, and responded.

To the extent that the agencies remained relatively upbeat during the 1920s, with upgrades usually outnumbering downgrades, only to revise their perceptions drastically after the crisis struck, questions about the stability of ratings through the cycle may be raised. Strikingly, we find that the procyclicality of ratings often emphasised in modern times is also a characteristic feature of their early experience, at a time when the agencies' business model (how they charged, how they were regulated, and how US regulators used them) was quite different from what it is now.⁵³ One interpretation is that the bond market collapse of 1931 was simply unpredictable. Although the agencies identified to some extent modestly increased risks in the second half of the 1920s, they were unable to forecast something of the magnitude of the global bond debacle of the 1930s. When it occurred, they ran for cover. Whether in so doing they simply reacted to the deterioration of financial conditions, or, as some commentators of modern episodes have suggested in different contexts, amplified the deterioration, is a matter for future research.

V. The performance of ratings during the interwar period: methodologies

The previous section offered an analysis of the reaction of rating agencies to the 1931 debt crisis. In the sections that follow, we take a closer look at the performance, or ratings accuracy, of the agencies. Accuracy is not the only performance criterion rating agencies care about, but it is definitely a major one. In the early 1930s, rating agencies found themselves under attack for their forecasting record. This was sufficiently serious to lead Standard Statistics to circulate globally a booklet meant to refute such allegations.⁵⁴ In this small pamphlet, Standard Statistics argued that it “did not grow to be the world largest statistical and investment advisory organization by giving inaccurate information and unsound advice. Rather, its growth may be attributed to the fact that it has conscientiously – and successfully – endeavoured to furnish reliable information and advice which, while by no means devoid of error, is right far more often than it is wrong”.⁵⁵

We assess performance via two principal methods.⁵⁶ The first focuses on default rates at different rating categories. Rating agencies indicate that relative rating accuracy is their primary objective. To be useful, ratings, at a minimum, should be able to provide a good ordering of risks among different credits. (This ability can be assessed by comparing either ex post default rates or recovery rates.) Since relative creditworthiness does not change dramatically over time, producing consistent rankings allows agencies to maintain ratings stability through the business cycle.⁵⁷

⁵³ In fact, the massive downgrades of September 1931 came only a few weeks after ratings were given “regulatory licence”. The extent to which this licence, and what it implied in terms of book-keeping (low-rated securities were penalised, thus encouraging banks to sell them in the open market), amplified the crisis by creating a feedback loop from downgrades to bond prices should be the topic of future research. For results pertaining to this point, see Gaillard (2008).

⁵⁴ Standard Statistics (1931). Harold (1934) is an early reference on contemporary interest in the performance.

⁵⁵ *Wall Street Journal*, 17 June 1931, Advertisement for Standard Statistics' 1931 brochure.

⁵⁶ See Cantor and Mann (2003) for a general discussion of performance measurement.

⁵⁷ Noting that many clients prefer ratings that are stable as well as accurate, Cantor and Mann (2007) explicitly analyse the trade-off between ratings accuracy and stability. We do not assess ratings stability in this study.

Thus, one indicator of the ability of rating agencies to order relative risks is the degree to which default rates tend to increase as ratings decline. The seminal work in the area is Hickman (1958, 1960), who examines the performance of American corporate bonds in 1900–43, and the related work by Atkinson (1967) and Altman (1989).⁵⁸ In general, research has documented a consistently negative association between default rates and credit rating categories (the lower the grade, the higher the default rate). Almost all post-war academic studies on the topic show this. In section VI, we examine whether ratings indeed sorted relative risks during the interwar era as well. We also examine whether ratings levels subsequently labelled by regulators as investment grade provided a significant degree of protection, in the form of much lower default rates.

However, documenting the simple existence of a negative relation between ratings and default rates is probably unsurprising and a fairly weak test of ratings reliability. Further, it does little to distinguish between the discriminatory power of one rating system versus another. A more precise assessment of a rating system is provided by the so-called cumulative accuracy profile (CAP), and the use of a statistic based on the CAP is the second method we pursue here. Like default rates, CAPs can be calculated for any cohort of securities at a single point in time for any given time horizon. As shown in the generic examples of Figure 6, the CAP curve indicates the share of total defaults among the borrowers as a function of the share of the lowest-rated borrowers in the sample: in some sense, they are a distant cousin of Lorenz curves that measure inequality in a given society. Namely, for any given share of the lowest-rated borrowers in the sample (eg the lowest 10% of the sample), the CAP curve charts what percent of defaults in the sample they account for (eg 50%). The more the CAP curve veers towards the upper left-hand corner close to the vertical and horizontal axes, the greater the proportion of defaults that occurs in the lowest rating categories.

This representation of ratings performance is independent of the particular scales used by agencies as well as the absolute level of default rates. Comparison then only requires using identical cohorts of securities across identical time periods. It is made by graphing different rating systems on the same chart and looking for the one that always provides a higher proportion of defaults for ratings in ascending order. While the concavity of Lorenz curves is bad news for a society's equality, the concavity of the CAP is good news for the "sorting capacities" of the rating system it measures: rating systems that produce "more concave" CAP curves are deemed superior to others. This is the case for rating system 1 in Figure 6, which dominates both rating systems 2 and 3. Of course, it can be the case that different curves provide ambiguous signals about their relative accuracy. This is when different parts of the curve lie closer to the upper left-hand corner, depending upon the portion of the rating scale that is examined. This is the situation for rating systems 2 and 3, which cannot be compared to one another.

It is possible to distil the information of any cumulative accuracy profile into a single statistic, the accuracy ratio (AR). This is the ratio of the area between the CAP curve and the 45 degree line, and the maximum possible area between the 45 degree line and a perfectly accurate rating system. (The means of calculation are discussed in the appendix.) The accuracy ratio has a number of attractive properties: a) it is always larger for any curve that lies above another; b) it effectively varies between -1 and 1 , much like a correlation measure, where 1 represents maximum accuracy when all defaulters are assigned the lowest ratings, and c) it is invariant to changes to the aggregate default rate that do not also

⁵⁸ Today, rating agencies regularly report evidence in their bond default studies, some of which look back as early as 1920 (Moody's (1997)). Moody's (2003), Standard & Poor's (2007), and Fitch (2007) provide extensions of default studies to sovereign ratings for the 1990s–2000s.

change the ratings distribution of the share of defaults.⁵⁹ In section VI, we compare the accuracy ratios of different rating systems.

VI. Rating categories and default rates

In this study, we assess accuracy by concentrating on defaults and default rates. Data for ratings are as described earlier.⁶⁰ Data for defaults come from *Moody's Manuals*, completed with information from the Annual Reports of the Foreign Bondholders Protective Council (FBPC).⁶¹ A natural starting point to the discussion of ratings reliability is the relevance of the letter-grade cut-off between investment grade (IG) and speculative grade (SG), although this cut-off did not yet exist (although it was already implied in the language used).⁶² Nonetheless, if we go by this rule, a broad consistency does emerge. We estimate that sovereign securities rated IG by Moody's exhibited a 4% average one-year default rate during 1928–37, compared to 18% for bonds in SG.⁶³ Results for other agencies are not dissimilar, pointing to a broad correspondence between rating categories and tendency to default similar to that found for corporate bonds (Hickman (1958)).

To delve further into the matter, Tables 5–7 provide a decomposition of default rates across time and agencies for three time horizons (1, 3 and 5 years). By and large, these tables are

⁵⁹ There are other schemes reducing the CAP to a single number with the above properties, but the key identifying assumption of the accuracy ratio of equation (1) is that the ratio, as calculated, weighs errors on the high end of the rating scheme equivalently to errors on the low end. Equivalently, the cost of having a defaulting borrower highly rated (type 1 error) is weighed equally to the cost having a non-defaulter rated too low (type 2 error). In terms of the intuition of the economic significance, the accuracy ratio can be viewed as a degree to which (measured in per cent) the rating system approaches the ideal in which all prospective defaulters were ex ante in the lowest category and all prospective non-defaulters were ex ante in the highest category. Namely, a zero accuracy ratio represents no relationship between ratings and defaults, while a ratio equal to 1, a perfect one. A 0.5 ratio means that, given randomly selected defaulter as well as non-defaulter, there is half again more probability of correctly inferring which is which based on their ratings than if the ratings were randomly assigned, ie the probability of correctly inferring the defaulter based on the rating is 75% (see Hanley and McNeil (1982) for the justification of this interpretation).

⁶⁰ When comparisons are made across agencies, we use the intersection between the four rating populations, meaning that it generally coincides with Fitch-rated issues.

⁶¹ In principle, actual recovery ratios ought to provide a good alternative, but available estimates of returns on defaulted bonds are too few to be really helpful. Eichengreen and Werley (1988) have computed internal rates of returns for a portfolio of 250 securities on the NYSE between 1920 and 1929. These estimates are the basis for the article by Eichengreen and Portes (1986) discussed earlier. However, of the securities analysed by Eichengreen and Werley, only 27 are, properly speaking, sovereign bonds (ie non-colonial, non-sub-sovereign, non-sovereign-guaranteed securities). This number is to be compared to the total population of 116 securities. Nonetheless, we performed some exercises using this limited population (not reported here). While the size of the sample does rule comparisons across agencies, we found broad consistency between ratings and rates of return, consistent with earlier findings in the literature. Using Moody's ratings, we found that speculative grade securities had an average rate of return of 2.86, while investment grade securities yielded 4.92 on average. Obviously, constructing additional estimates for recovery rates would be a daunting task.

⁶² As stated above, it emerged during the 1930s when regulators started associating "speculative" securities with grades below the four top notches.

⁶³ Authors' computations. The standard test for the difference in means of these two default rates based on sample sizes produces a test-statistic of 4.8, well above standard levels of statistical significance. The difference in default rates remains striking if we include the BBB category in the speculative grade category: for instance, while the absolute difference in default rates is slightly smaller at 10%, the test-statistic rises to 5.9. Looking at sovereign securities to examine differences in default rates is particularly informative because of the extent of the shock. By means of comparison, average default rates for "investment grade" and "speculative grade" corporate securities for 1930–34 were 0.56% and 8.42%, respectively (Source: Moody's (2009))

all consistent with default rates rising as ratings decline for most years before and during the financial crisis. Beyond this, however, the interwar experience stands out as exceptional in several respects.⁶⁴

First, considerable variability is present in default rates at various time horizons: For example, the one-year high-grade default rate for Fitch, which averages 3% in the years 1928 to 1937, ranges between 0 and 16% for individual years; at the longer horizon of five years, the high-grade default rate for Moody's ranges between 2 and 32%. Some variability is consistent with more recent evidence (Cantor and Packer (1994)). However, the numbers involved are much greater than those documented more recently, probably indicative of the degree to which the foreign government debt crisis and the Great Depression were unanticipated and certainly beyond cyclical norms.⁶⁵

Second, the numbers show high rates of casualties *even for high-grade credits*: nearly an order of magnitude difference at all horizons relative to recent experience. For instance, the *average* high-grade one-year default rate (for the years 1928–37) of between 2 and 4% for the four rating agencies (Table 5) dwarfs Moody's *worst* single-year default rate for investment grade credit during 1983–2001 of 0.5%. Similarly, high-grade default rates at the three-year horizon (for the years 1928–37) average between 8 and 13% for the four rating agencies (Table 6), compared to Moody's worst year of 2% in the sample cited above.⁶⁶

Third, it is worth noting that rating agencies differ quite a bit among themselves in the interwar era. Moody's average one-year default rate for BBB (Baa) issues of 10% corresponds to default rates for Fitch, Poor's and Standard Statistics of 4, 5 and 4%, respectively. For average three-year BBB (Baa) default rates, Moody's 28% compares to 18%, 17% and 16% for Fitch, Poor's and Standard Statistics.⁶⁷ The differential may capture different levels of performance, but it could also be indicative of the fact that before the correspondence of ratings implied by regulation was well-established, the default rates could well have been expected to differ at seemingly similar rating categories. For instance, Moody's may have thought at that time of their Baa rating as corresponding to a higher default rate than the other rating agencies thought of their BBB-equivalent.⁶⁸ Another example is provided by Standard Statistics' higher-grade rating default rates, which were consistently lower than those of the other three agencies at all time horizons, suggesting

⁶⁴ The default rates and accuracy ratios reported in the following two sections and in Tables 4–7 were also calculated for a smaller issuer-basis sample, whereby only one bond (either the most junior or the one for which the most data were available) was taken into the sample per country. None of the results reported in the following two sections differed significantly for the smaller sample.

⁶⁵ Of course, countries today have the recourse of going to the IMF when they need funding, while in the interwar period default was often the only solution to restructure external claims. An interesting exercise would be to compare default numbers for the interwar period with the sum of defaulters and those receiving IMF support in the modern period.

⁶⁶ To be sure, some of the difference in default rates compared with more recent experience is due to the lower frequency at which observed ratings were available. However, even adding one year to the time horizon yields average default rates for the more recent period that are much lower than those calculated for the interwar era.

⁶⁷ The test statistics for the difference of the one-year Baa default rates between Moody's and the other agencies are 2.17 (Fitch), 2.13 (Poor's) and 2.01 (Standard), all above standard levels of statistical significance. Though the absolute differences in three-year default rates are greater in magnitude, there are not enough independent observations (the rates must be taken three years apart) to achieve statistical significance.

⁶⁸ In the post-war era; if anything, evidence suggests that Moody's ratings probably corresponded to a stricter scale than those of some other agencies (Cantor and Packer (1997)).

either a different quality cut-off for high-grade credits (in conflict with what external observers perceived) or a more pessimistic outlook than the other agencies.⁶⁹

Fourth, these differences are most noticeable in the default rates where the base year is at the onset of the debt crisis. If we go by the IG/SG cut-off, we see that several rating agencies had initially high casualty rates in the investment grade category. Fitch is a good illustration, with a 16% default rate (for a one-year-ahead horizon) in 1931. This may have accounted for its particularly aggressive downgrades during that year and the following one, suggesting that the crisis may have contributed to greater homogeneity in rating agencies' scales. In any case, this reinforces the notion that the extent of the crisis was generally not anticipated by the agencies.

Figures 7 and 8 provide further insights. Figure 7 shows the composition of the (last observed) ratings for the four rating agencies at the time of default for the 44 defaults in the sample.⁷⁰ Defaults surged across many rating classes. "High-grade" ratings appear to have provided little guarantee against default during the financial crisis. More than one-half of all defaulted bonds are observed to have a high-grade rating (the equivalent of Moody's Baa or higher) from at least one agency. In the case of Fitch and Poor's, 39% of the last observed ratings of the defaulting issues are high-grade. In the case of Moody's, fully 45% were rated high-grade. Standard Statistics has "only" 25% of the defaulting issues rated high-grade.⁷¹ The incidence of high-grade defaults at the time was not merely due to bonds at the fourth highest letter grade level. A good chunk of the high-grade defaults occurred with observed ratings even higher than the fourth highest letter grade level: defaults of A grade (or equivalent) or higher accounted for up to 23% (Poor's) of the high-grade defaults.⁷² Even starker results are obtained when we move a bit further back in time and check the last observed ratings of defaulted issues one year ahead of default (Figure 8). Between 39% (Standard Statistics) and 66% (Moody's) of the defaulting bonds are observed to have a

⁶⁹ This finding is not inconsistent with our earlier finding that in 1931 Standard Statistics did not downgrade government bonds as aggressively as other agencies. The test statistic for the difference of the one-year investment grade default rates of Standard Statistics and all of the other agencies taken together is 2.0. Relative to Moody's alone, the test statistic for the difference of the one-year IG default rate of Standard Statistics is 2.3. While the differential lags at which we observe ratings during the pre-war era probably account for some of the higher Moody's Baa default rates and lower Standard Statistics (overall) default rates, examination of default rates at marginally shorter horizons for Moody's and marginally longer horizons for Standard Statistics indicates that the lags cannot account for most of the difference.

⁷⁰ Computations were performed using the latest available rating. The fact that the ratings were based on snapshots taken at different months of the year implies that the ratings used of some agencies for this study might have had an informational advantage relative to others by being more recent. We checked whether the timing of defaults relative to observed ratings might have biased the comparison of the agencies with regard to ratings at default, eg accounting for Standard Statistics having a lower percentage of highly rated issues at default. However, the average number of months elapsed between the latest available rating and the dates of defaulted issues is quite similar: 4.9 for Fitch, 5.4 for Poor's, 6.2 for Standard Statistics and 6.8 for Moody's. If anything, Standard Statistics' ratings were at a slight disadvantage.

⁷¹ This contrasts notably with the general default experience of the last two decades of the 20th century. Cantor and Mann (2003) state that the average Moody's rating for defaulting issues during 1983–2001 is CCC. Of course, these ratings from the modern era were continuously updated, giving them an informational advantage than our interwar ratings. However, ratings of issues that default at longer horizons (where the informational advantage is lost) are also significantly lower in the modern than in the pre-war era.

⁷² No bond had the top rating (AAA/Aaa/A**/A1+) at the time of default. The one bond which had top ratings at some time prior to default was the 1924 German bond, rated AAA by Fitch four years before its default in 1934. Two Dominican Republic securities had ratings in the second highest category at the time of default (rated AA by Fitch). Several bonds were rated in this category at some point before their default (Dominican bonds: Fitch and Standard Statistics), Germany 1924 (Moody's), Cuba 1930 (Poor's), Panama 1928 (Fitch, Poor's, Standard Statistics). Many bonds were rated A by the agencies at the time of default or a few years before.

high-grade rating one-year ahead of default.⁷³ Again, there are many future defaulters in our sample rated A grade (or equivalent) or above one year ahead of time (more than half the high-grade defaulters).

The results may be interpreted as supporting the view that, although rating agency observers by and large were able to rank relative risks, they underestimated the extent of the interwar debt crisis. This may indicate something about the agencies' predictive capacity, the nature of the crisis, or both.

VII. Performance and accuracy ratios

We suggested that the use of absolute scales or the IG/SG cut-off, which only developed during the 1930s, may be anachronistic. In what follows, we compare the performance of the four agencies' separate ratings, and an average of their ratings, using the accuracy ratios discussed above. As already indicated, these ratios have the attractive property of focusing solely on relative grades, and thus permit us to avoid any assumptions regarding the correspondence of rating scales across agencies.

a) The relative performance of agencies

The need to obtain credit ratings as well as matching market yields across the four agencies produces a set of more than 100 bonds, though only a subset of these are used for any single cohort since not all bonds are rated by the same agencies, and they enter and exit the sample due to new issuance, redemption, default, etc (for instance, 89 bonds are used to assess Moody's accuracy ratios in 1931). As before, we consider three horizons (one, three and five years). To capture the performance over the wave of interwar defaults and downgrades, we examine the one-year-ahead predictive performance of the various ratings between 1931 and 1936 (since there were no defaults in 1935, one-year accuracy ratios cannot be calculated for that year). We look at the three-year-ahead performance for the (non-overlapping) cohorts of 1929, 1932 and 1935. And, finally, the five-year-ahead performance is examined for the non-overlapping cohorts of 1929 and 1934.

The results (Table 8) suggest that, quite apart from the issues discussed above of the respective scales used by the agencies, there were on occasion differences in accuracy in the ranking of relative credits. These differences come into clearest relief when we differentiate between the performance of ratings over the short-term 1-year horizon, and those over the longer 3- to 5-year horizons.

Over the 1-year horizon, Fitch, Moody's and Standard Statistics each belong to two cohorts in which they have the highest accuracy ratio (AR), while Poor's has the lowest average mean and median ratios over the period. Though Fitch's AR was low in the first year, with the poorest forecast of 1931 defaults, it improved considerably in the years that followed, probably reflecting the many rather rapid rating changes it made subsequently. Tests of statistical significance of these differences lend some credit to Moody's dominance.⁷⁴

⁷³ This compares to an average Moody's rating one year prior to default of B in the 1983–2001 sample mentioned above.

⁷⁴ In a series of articles, Hanley and McNeil (1982, 1983) derive test statistics for the significance of differences in the accuracy ratios of the two rating systems. Though the null hypothesis of similar accuracy ratios cannot be rejected at conventional levels of significance for most of the pair-wise comparisons in each year of the 1-year horizon, Moody's accuracy ratio is significantly higher than both Fitch's and Poor's (at the 95% level of statistical significance) over the one-year horizon for 1931. Standard Statistics' ratio is significantly higher than

Over the longer-term 3- and 5-year horizons, Moody's or Standard Statistics have the highest accuracy ratios among the 4 agencies in 60% and 40% of the cohorts examined, respectively, and the two agencies have the highest mean and median accuracy ratios at these time horizons as well.⁷⁵ The clear differentiation in longer-term performance is interesting *per se*, and may be related to the evidence presented above that Moody's was the market leader in ratings, while Standard Statistics was actively raising its profile. Current supporters of the view that competition among agencies may promote the accuracy of their ratings can find comfort in this result.⁷⁶

In an indication that the individual rating agencies might be getting separate, valuable, signals about the creditworthiness of sovereign borrowers, the accuracy ratios of the consolidated ratings (average rating) are often superior to those of each of the single agencies at longer horizons. The average rating scored higher on average than all of the individual agency ratings at the three- and five-year horizons, with a mean accuracy ratio higher than each of the single agencies.⁷⁷ In six of the ten cohorts, the accuracy ratio of the average ratio was significantly higher (at the 95% level of statistical significance) than at least one of the rating agencies.

b) Agencies vs the "market"

Finally, we compare the agencies with a measure of "market" ratings. To compute market ratings, we use a method detailed in Breger et al (2003) to transform yield spreads (collected using prices in the *Wall Street Journal*)⁷⁸ into ratings. Of course, to the extent that ratings influence markets (and perhaps conversely, too) the information in ratings may not be different from that in yield-implied ratings. Yet this is precisely what we seek to discuss: we found in the previous section that rating agencies were quite able to get relative risks right but missed the risks of an aggregate shock. One question that remains to be addressed is whether this particular failure reflected a worse, similar, or even better performance than that of the market. Since this was the time when rating agencies started to be relied upon by regulatory authorities in the US, there is a possibility that the agencies had somehow outsmarted the market – even if only by a modest margin.

Poor's in 1931 and Fitch's in 1934. The performance of Standard Statistics probably reflects, at least in part, the shorter lag (than other agencies) between the previous year's observed rating for which rating cohorts are determined and the periods over which defaults are observed. This factor will be most important for the 1-year accuracy ratios. By contrast, the performance of Moody's ratings is particularly striking given that its observed ratings have the longest lag.

⁷⁵ In terms of significance tests, Moody's ratio is significantly higher (at the 95% level of statistical significance) than both Fitch's and Poor's in 1929 and 1932 over the 3-year horizon. Mean and median accuracy ratios were calculated over all available yearly cohorts, and thus contain overlapping observations at horizons greater than one year.

⁷⁶ Competition for accuracy ought to be differentiated from competition for customers, which is said to encourage rating inflation and a deterioration in performance. Given the difference between the interwar model, whereby agencies charged their readers, and the current one, whereby they mostly charge issuers, prevent us from making any inference on this subject.

⁷⁷ We also checked whether accuracy was improved by using the higher or lower rating, but the average was the aggregator that resulted in the highest accuracy ratios. Measuring the extent to which the accuracy would have been improved by using variables simultaneously, such as the actual ratings and yield-based ratings, or the level of ratings and the degree of disagreement between the ratings, would be a good subject for future research.

⁷⁸ While the prices taken from the *Wall Street Journal* were indicative rather than transactional prices, issues such as non-trading and staleness of quotes that could significantly bias the results in other empirical exercises are unlikely to be important in this paper given the relatively low frequency of the analysis (annual) and the fact that the market prices were generally observed at a later point in the period than ratings.

Table 9 reports the difference between the accuracy ratios of the agency ratings and synthetic ratings derived from market prices for the same cohorts. Therefore, a negative number in any cell indicates that the agency ratings had lower accuracy ratios than the yield-based metrics and thus were not as effective in sorting risks. The results suggest that market-based ratings modestly outperform observed agency ratings at the one- and three-year horizons, having a higher AR in about 16 of the 25 cases (4 agency ratings plus average rating over five separate years) at the one-year horizon, and a higher AR in two-thirds of the cases at the three-year horizon. At the longer 5-year horizon, actual agency ratings appear to outperform the yield-implied ratings (YIRs), though only modestly.⁷⁹

That said, the difference in the accuracy ratios between observed ratings and yield-implied ratings at all maturities, while large for specific years and agencies, is relatively small when the mean is taken across all the cohort years. In addition, tests of statistical significance are unable to reject the null hypothesis of no difference between agency and market-based ratings in 15 out of 25 cases.⁸⁰ In sum, there does not seem to be anything specific or exceptional about the performance of rating agencies in assessing borrowing governments' relative credit risk over the interwar period compared to what could have been inferred from market prices, except at longer time horizons.⁸¹ This, we reckon, is certainly consistent with the fact that, even if they do not face alleged conflicts of interest, rating agencies should not be expected to stand too far outside market forecasts. And if the market at large missed the crisis, why should rating agencies have been any different?

Conclusions

The interwar period offers a rich environment where we can examine the record of rating agencies. While many aspects of ratings could be considered, this paper has focused on the ratings of foreign debt issued on the NYSE. Our main empirical finding is that the performance of rating agencies during the interwar era was not exceptional. Ratings seem to have reacted to similar indicators as they do in modern studies. Ratings above subsequent regulatory cut-offs provided little guarantee against default during the financial crisis. Rating agencies generally did not exhibit forecasting capacities superior to those embedded in available market prices. While this probably means that the agencies, like many other players in the financial industry, did not see the interwar debacle coming, it is also consistent with the view that their forecasting capacities at the time were not particularly outstanding.

⁷⁹ Again, the greater lag at which our agency ratings are observed relative to market yields (greatest for Moody's, least for Standard Statistics) might account for some of the (marginally) superior market yield performance, particularly at short horizons. As reported in Table 7, though the accuracy ratio tends to be higher for actual agency ratings than for yield-based ratings for the two selected cohorts at the five-year horizon, the mean difference calculated across all annual cohorts find agency ratings to (slightly) outperform market yields only in the case of Moody's. The improvement in the ratings performance relative to the YIRs at longer maturities parallels the results for more recent samples using Moody's ratings (Cantor and Mann (2003)).

⁸⁰ Yield-implied ratings have significantly higher ARs (at the 5% level of statistical significance) at the one-year horizon in 1931 for Fitch, Poor's, Standard Statistics and the average rating; at the one-year horizon in 1933 for Moody's; at the three-year horizon in 1932 for Fitch, Poor's and Standard Statistics. Agency ratings have significantly higher ARs than yield-implied ratings in 1929 at both the 3-year and 5-year horizons for Moody's.

⁸¹ To give a sense of the economic significance of one of the largest differences in Table 9, the fact that Moody's accuracy ratio in 1929 at the 5-year horizon was 25% greater than that of the yield-implied ratings implied that had one defaulting and non-defaulting bond each been chosen from that cohort, the Moody's rating would have had a 12.5% greater probability of correctly identifying the defaulter than the yield-implied rating (25% of 50%).

At the same time, we also reported some evidence of rating agency insights into the crisis. The decline in the quality of securities issued in the 1920s, for instance, did not go unnoticed, although its extent was vastly underestimated *ex post*. Similarly, we found that market discipline may have operated in the shape of reputational brand concerns, which may explain the greater coverage and marginally superior forecasts by certain rating agencies, such as Moody's and Standard Statistics, the two market leaders.⁸²

These findings have relevance for students of the Great Depression. They suggest that, despite what representatives of the rating industry such as Max Winkler later claimed, the agencies were not really the Cassandra of the international debt crisis. There is little evidence that the "unethical" behaviour of bankers could have been reined in had the rating agencies been given a bigger say. The conclusion may well be that, as some earlier characterisations of the crisis emphasised, the key drivers of the crisis were largely unpredicted developments such as the decline in commodity prices that undermined the fiscal base of many borrowers and triggered an extraordinary crop of failures.

This conclusion provides food for thought. Given the fact that the performance of agency ratings did not shine at the onset of the sovereign debt crisis, it is somewhat strange that it was precisely at this time that regulators started relying on them. The initial decision by the OCC to rely on ratings appears to have been taken in the late summer of 1931, when defaults were still few (ironically, a few weeks before Moody's massive downgrades), but it was sustained over the next few years at a time when it should have been clear that reliance on ratings would have provided limited protection against the crisis. While the behaviour of rating agencies during the foreign debt crisis of the interwar period was typical of other market monitors (ie delay in recognising the severity of the crisis, then massive downgrades), the reaction of policymakers was nonetheless to reach out and enshrine their credit ratings in the new regulatory frameworks they were creating. This reaction was exactly the opposite of the one many scholars now advocate. Future research will have to explain why this happened.

References

- Afonso, A. (2003), "Understanding the Determinants of Sovereign Debt Ratings: Evidence for the Two Leading Agencies", *Journal of Finance and Economics*, Vol. 27, No.1.
- Altman, Edward (1989), "Measuring Corporate Bond Mortality and Performance", *Journal of Finance*, Vol.44.
- Atkinson, Thomas R. (1967), *Trends in Corporate Bond Quality*, NBER.
- Babson, Robert W. (1910) "Sources of Market News"; *The ANNALS of the American Academy of Political and Social Science*, Vol. 35; 135, pp. 617-628.
- Benmelech, Efraim and Jennifer Dlugosz, 2009. "The Credit Rating Crisis," NBER Working Papers 15045.
- Beghoff, Helmut, "Civilizing capitalism? The beginnings of Credit Rating in the United States and Germany", Inaugural Lecture at the German Historical Institute, Washington, Nov 14, 2008.
- Bhatia A. (2002), "Sovereign Credit Ratings Methodology: An Evaluation", IMF Working Paper, WP/02/170.

⁸² In 1941, Standard Statistics and Poor's would merge to create Standard & Poor's.

- Board of Governors of the Federal Reserve System (1943), *Banking and Monetary Statistics 1914-1941*, Washington, D.C.
- Breger, Ludovic, Lisa Goldberg, and Oren Cheyette (2003), "Market Implied Ratings", *Barra Credit Series, Research Insights*.
- Cantor, Richard and Christopher Mann (2003), "Measuring The Performance of Corporate Bond Ratings", *Moody's Special Comment*.
- Cantor, Richard and Christopher Mann (2007), "Analyzing the Tradeoff between Ratings Accuracy and Stability", *Journal of Fixed Income*, Spring.
- Cantor, Richard and Frank Packer (1994), "The Credit Rating Industry", *Federal Reserve Bank of New York Quarterly Review*, Summer-Fall.
- Cantor, Richard Martin and Packer, Frank (1996), "Determinants and Impact of Sovereign Credit Ratings," *Economic Policy Review*, October.
- Cantor, Richard and Frank Packer (1997), "Differences of Opinion and Selection Bias in the Credit Rating Industry", *Journal of Banking and Finance*, Vol.21.
- Carruthers, B. and Barry Cohen (2006) "The Mechanization of Trust: Credit Rating in 19th-c. America." Working paper.
- Chandler, Alfred (1956), *Henry Varnum Poor: Business editor, analyst, and reformer*, Cambridge, Mass.: Harvard University Press.
- Committee on the Global Financial System (2005), "The role of ratings in structured finance: Issues and implications", *CGFS Publications*, No. 23, Bank for International Settlements, Basel.
- Committee on the Global Financial System (2008), "Ratings in structured finance: What went wrong and what can be done to address shortcomings", *CGFS Publications*, No. 32, Bank for International Settlements, Basel..
- Contreras, Manuel (1990), "Debt, Taxes, and War: The Political Economy of Bolivia, c. 1920-1935", *Journal of Latin American Studies*, Vol.22.
- Dominguez, K.M., R.C. Fair, M.D. Shapiro (1988), "Forecasting the Depression: Harvard versus Yale", *American Economic Review*, Vol.78, 4, pp. 595-612.
- Edwards, George (1928), "Government Control of Foreign Investments", *American Economic Review*, Vol.18, No.4.
- Eichengreen, Barry, and Peter Lindert (1989), *The International Debt Crisis in Historical Perspective*, MIT Press, Cambridge.
- Eichengreen, Barry and Richard Portes (1986) "Debt and default in the 1930s. Causes and consequences", *European Economic Review*, Vol.30, pp. 599-640.
- Eichengreen, Barry, and Jeffrey Sachs (1985), "Exchange Rates and Economic Recovery in the 1930s," *Journal of Economic History* 45, pp. 925-946.
- Eichengreen, Barry, and Werley, Carolyn (1988), "How the Bondholders Fared: Realized Rates of Return on Foreign Dollar Bonds Floated in the 1920s", Working Paper No.8869, University of California, Berkeley.
- Favero, Giovanni (2007), "Weather forecast or rain-dance? On inter-war business barometers", *Working Paper*, University of Venice.
- Federal Reserve Committee on Branch, Group, and Chain Banking, 1932, 225 bank suspensions: Case histories from examiners reports

Ferri, Giovanni, L.-G. Liu, and Joseph Stiglitz (1999), "The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis", *Economic Notes, Banca Monte dei Paschi di Siena SpA*, No.3.

Fitch (2007), *Sovereign Rating Transition and Default Study 1995-2006*, February 7.

Fitch (various years), *Fitch Bond Books*.

Flandreau, Marc (2003), "Caveat Emptor – Coping with Sovereign Risk Under the International Gold Standard, 1871-1913", in Flandreau, Marc, C-L. Holtfrerich, and Harold James (eds.), *International Financial History in the Twentieth century. System and anarchy*, Cambridge University Press.

Gaillard, Norbert (2008), *Les méthodologies de notation souveraine*, Unpub. Diss., Institut d'Etudes Politiques de Paris.

Grieser (1938), *The British investor and his sources of information*, unpublished dissertation.

Hanley, James and Barbara McNeil (1982). "The Meaning and Use of the Area under a Receiver Operating Characteristic Curve" *Radiology*, Vol. 134, No. 1, pp. 29-36.

Hanley, James and Barbara McNeil (1983). "A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases" *Radiology*, vol. 148, No. 3, pp. 839-843.

Harold, Gilbert (1934), "Do bond ratings forecast the market?" *Barron's*, Vol.14, No.53.

Harold, Gilbert (1938), *Bond Ratings as an Investment Guide*, Ronald Press Company, New York.

Hickman, W. Braddock (1953), *The Volume of Corporate Bond Financing Since 1900*, Princeton University Press.

Hickman, W. Braddock (1958), *Corporate Bond Quality and Investor Experience*, Princeton University Press.

Hickman, W. Braddock (1960), *Statistical Measures of Corporate Bond Financing since 1900*, Princeton University Press.

James, Harold (2001), *The End of Globalization: Lessons from the Great Depression*, Harvard University Press.

Jüttner J.D. and McCarthy J. (2000), "Modelling a Rating Crisis", Macquarie University, Unpublished manuscript, Sydney.

Klug, Adam, John S. Landon-Lane, and Eugene N. White, 2005. "How could everyone have been so wrong? Forecasting the Great Depression with the railroads," *Explorations in Economic History*, Elsevier, vol. 42(1), pages 27-55, January.

Lamar, W. R., 1907, *Investments: What and When to Buy. The Use of Statistics in Accumulating Wealth*, Mass.: The Babson System, Wellesley Hills.

Lewis, Cleona (1938), *America's Stake in International Investments*, Brookings Institution, Washington.

Madden, John, Nadler, Marcus and Sauvain, Harry (1937), *America's Experience as a Creditor Nation*, Prentice-Hall Inc, New York.

Mintz, Ilse (1951), *Deterioration in the Quality of Foreign Bonds Issued in the United States 1920-30*, NBER, Cambridge.

Moody's Investors Service (1919), "The Credit of Foreign Governments", *Moody's Investment Letter*, April 3.

Moody's Investors Service (1924), *Foreign and American Government Securities*.

- Moody's Investors Service (1995), *Corporate Bond Defaults and Default Rates 1970-1994*.
- Moody's Investors Service (1997), *Historical Default Rates of Corporate Bond Issuers, 1920-1996*.
- Moody's Investors Service (2003), *Sovereign Bond Defaults, Rating Transitions and Recoveries (1985-2002)*, February.
- Moody's Investors Service (2007), *Sovereign Defaults and Recovery Rates, 1983-2006*, June.
- Moody's Investors Service (2009), *Corporate Default and Recovery Rates, 1920-2008*, Special Comment, February.
- Moody's Investors Service (various years), *Moody's Manuals*.
- Morton, Walter A. (1939), "Liquidity and Solvency", *American Economic Review*, Vol. 29, No.2.
- Nash, Robert Lucas (1889), *Fenn's Compendium of the English and Foreign Funds*, London: Effingham Wilson.
- Norris, James, 1978, *R G Dun & Co, 1841-1900: The Development of Credit Reporting in the Nineteenth Century*, Westport: Greenwood Press.
- Partnoy, Frank (2006), "How and Why Credit Rating Agencies Are Not Like Other Gatekeepers", in Yasuyuki Fuchita, and Robert E. Litan (eds.), *Financial Gatekeepers: Can they Protect Investors?*, Brookings Institution Press and the Nomura Institute of Capital Markets Research.
- Partnoy, Frank (2009), "Historical Perspectives on the Financial Crisis: Ivar Kreuger, the Credit-Rating Agencies, and Two Theories About the Function, and Dysfunction, of Markets", *Yale Journal on Regulation*, Vol. 26.
- Poor's (various years), *Poor's Volumes*.
- Reinhart, Carmen M. (2002), "Sovereign Credit Ratings Before and After Financial Crises," in *Ratings, Rating Agencies and the Global Financial System*, edited by Levich R., Majnoni G. and Reinhart C., Kluwer Academic Publishers, Boston.
- Reisen, Helmut and Julia Von Maltzan (1999), "Boom and Bust and Sovereign Ratings", *OECD Development Centre, Working Papers No.148*.
- Rippy, J. Fred (1950), "A Bond-Selling Extravaganza of the 1920's", *Journal of Business of the University of Chicago*, Vol.23, No.4.
- Sandage, Scott (2005), *Born Losers: A History of Failure in America*. Cambridge: Harvard University Press.
- Standard Statistics Company (various years), *Standard Bond Descriptions*.
- Standard & Poor's (2007), *Sovereign Defaults And Rating Transition Data: 2006 Update*, February.
- Standard Statistics (1931), *A Forecaster's view of forecasting*, New York: Standard Statistics.
- Sy, Amadou, (2004) "Rating the rating agencies: Anticipating currency crises or debt crises?", *Journal of Banking & Finance*, Volume 28, Issue 11, November, pp. 2845-2867.
- Sylla, Richard (2002), "An Historical Primer on the Business of Credit Rating", in *Ratings, Rating Agencies and the Global Financial System*, edited by Levich R., Majnoni G. and Reinhart C., Kluwer Academic Publishers, Boston.
- Tomz, Michael (2007), *Reputation and International Cooperation*, Princeton University Press.

White, Eugene N. (2009), "Lessons from the history of bank examination and supervision in the United States 1863-2008", paper prepared for the "Conference on Financial Market Regulation after Financial Crises: The Historical Experience".

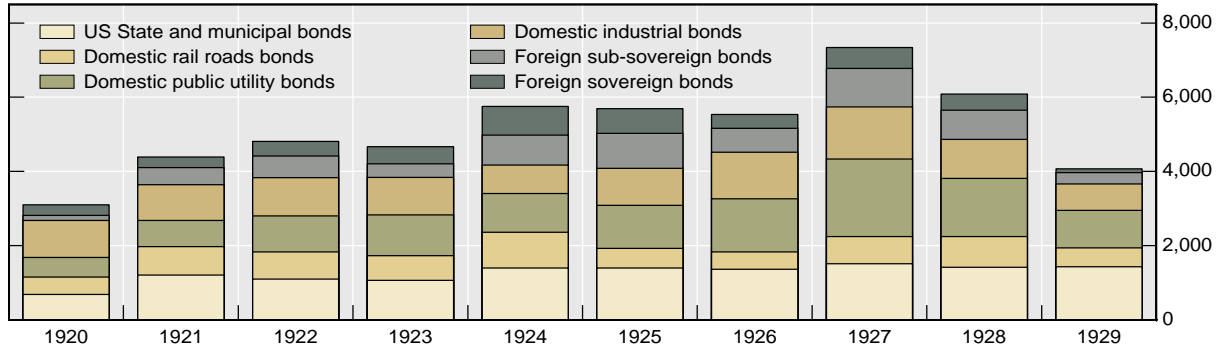
Wigmore, Barry (1985), *The Crash and its Aftermath: A History of Securities Markets in the United States*, New York: Greenwood Press.

Winkler, Max (1933), *Foreign Bonds – An Autopsy*, New York: Roland Swain Company.

Young, Ralph (1930), *Handbook on American Underwriting of Foreign Securities*, U.S. Department of Commerce, Washington, D.C.

Figure 1
Bond issuance on the NYSE (1920–29)

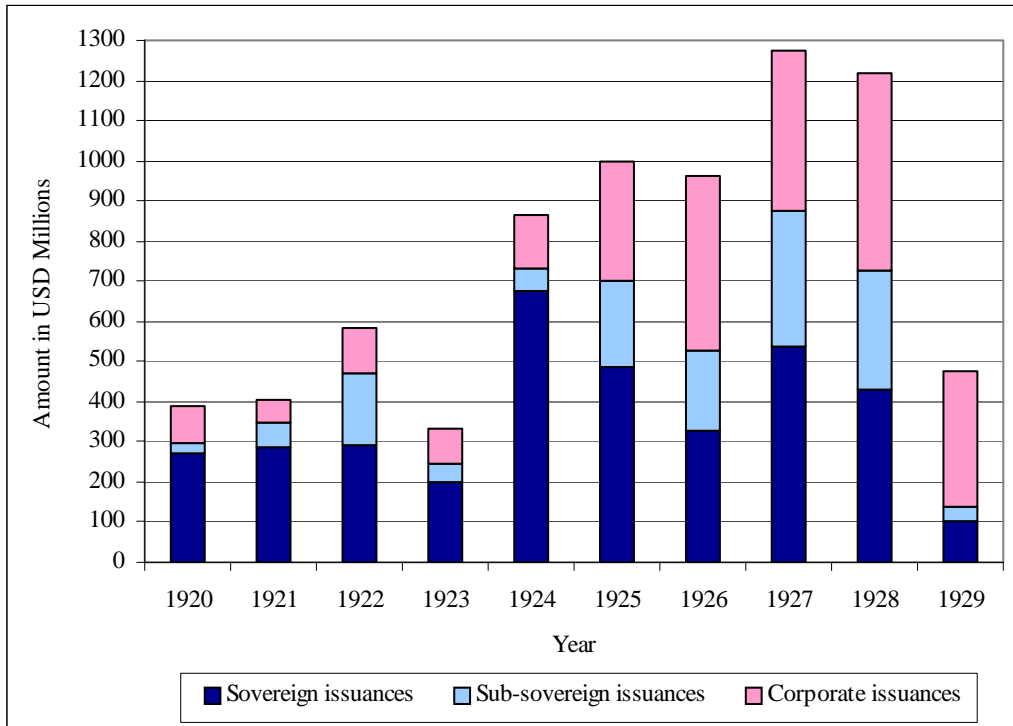
In millions of US dollars



Source: Hickman (1953), pp 253–5, for domestic railroad, public utility, and industrial figures. Board of Governors of the Federal Reserve System (1943), p 487, for US state and municipal figures. Authors' computations from *Moody's Manuals* for foreign sub-sovereign and sovereign figures.

Figure 2.a

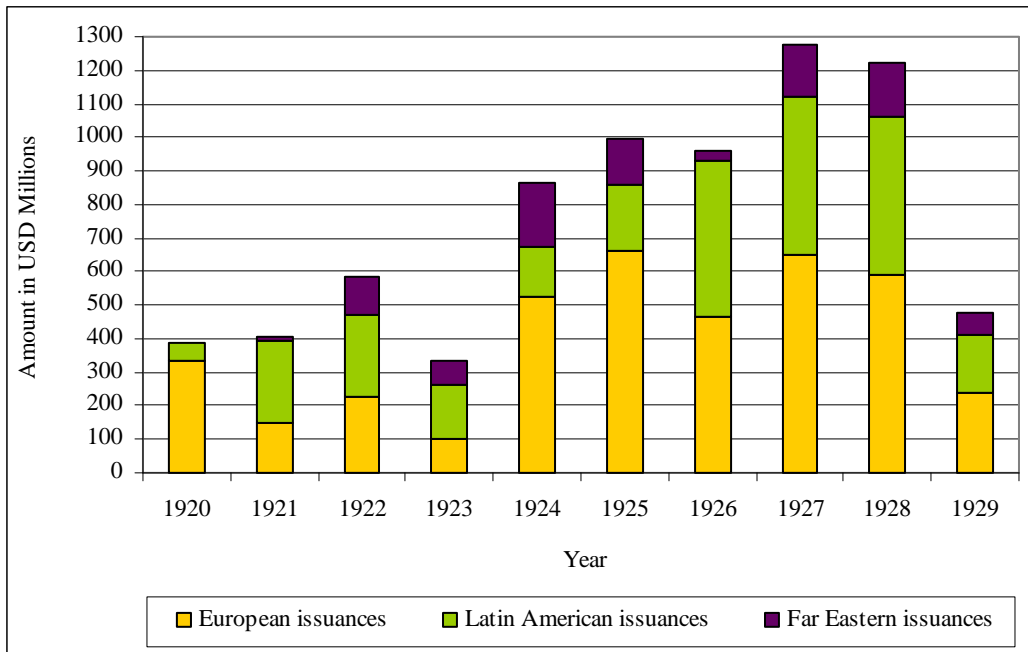
Foreign bonds: sovereign, sub-sovereign and corporate bonds
(issues on the NYSE, 1920–29, USD millions)



Sources: authors' computations from Young (1930).

Figure 2.b

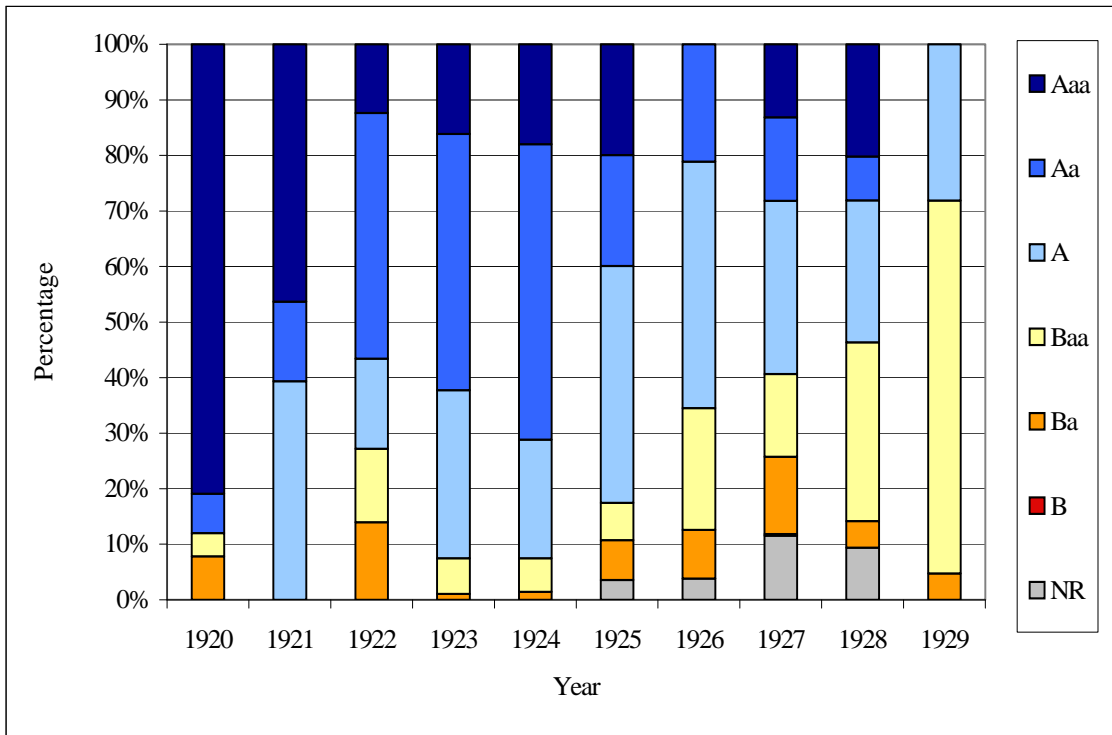
Total European, Latin American and Far Eastern bonds
(issued on the NYSE, 1920–29)



Sources: Authors' computations from Young (1930).

Figure 3

Ratings assigned by Moody's to bonds in sample by year of issuance

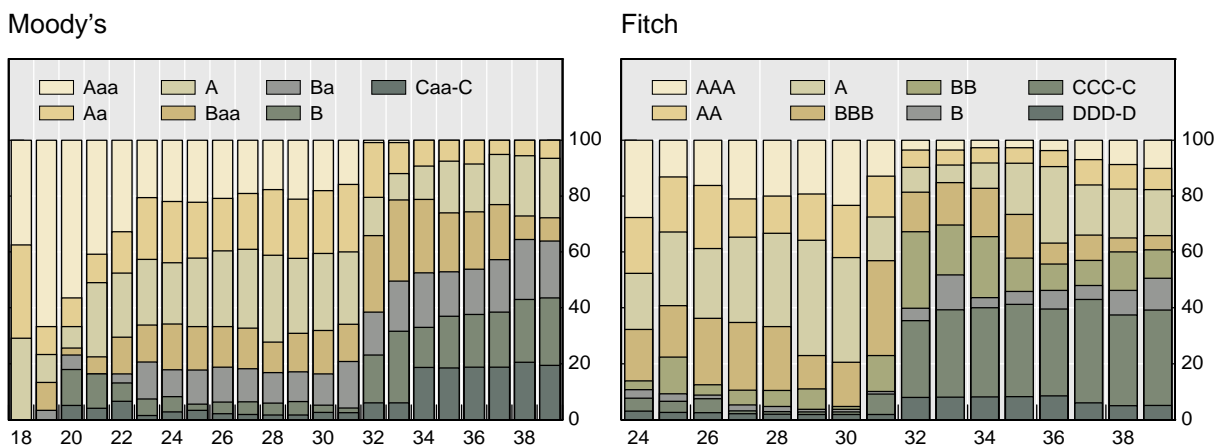


Sources: Authors' computations from *Moody's Manuals* (various years) and Young (1930). We consider sovereign, sub-sovereign, and state-guaranteed. NR denotes no rating assigned by Moody's for policy reasons. We looked in manuals for first available grade for each security issued in the stated year.

Figure 4

Composition of sovereign bonds rated in Moody's and Fitch Manuals (1918–37)

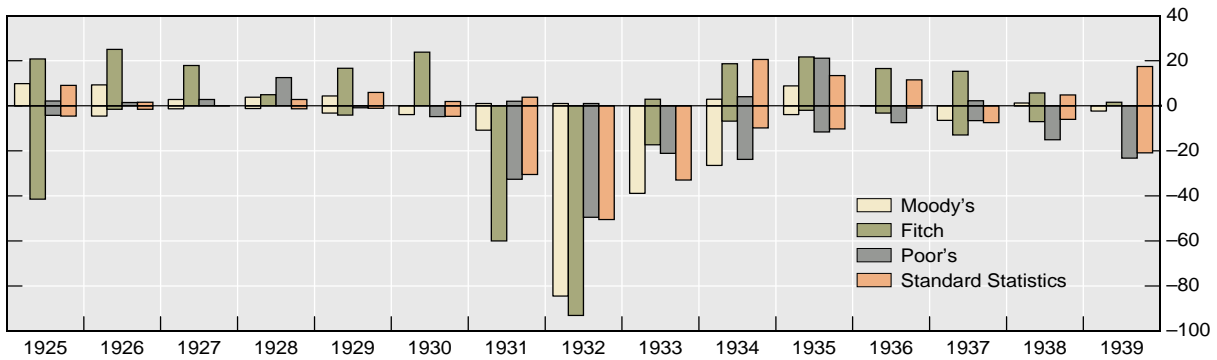
In per cent



Source: Gaillard (2008).

Note: Grades are from the volume of stated year. Moody's was released in January, Fitch in August.

Figure 5
Upgrades and downgrades as fraction of outstanding ratings¹
 In per cent

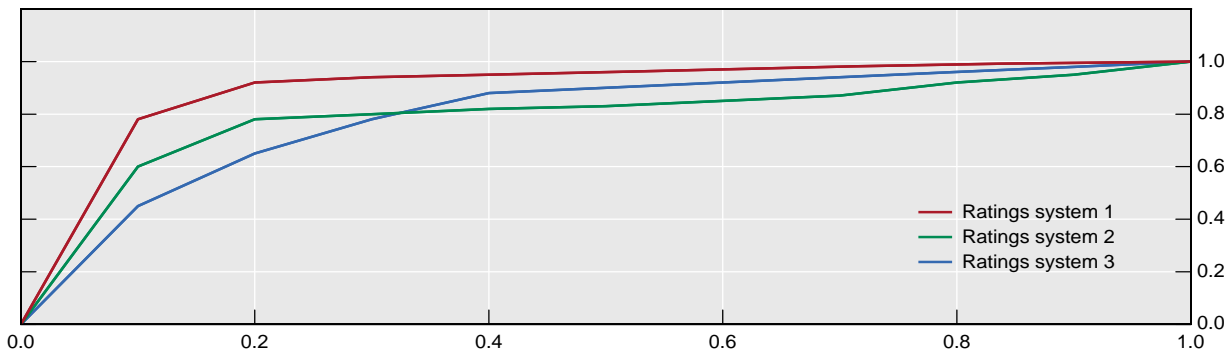


Positive bars indicate upgrades, negative indicate downgrades.

Source: Authors' computations.

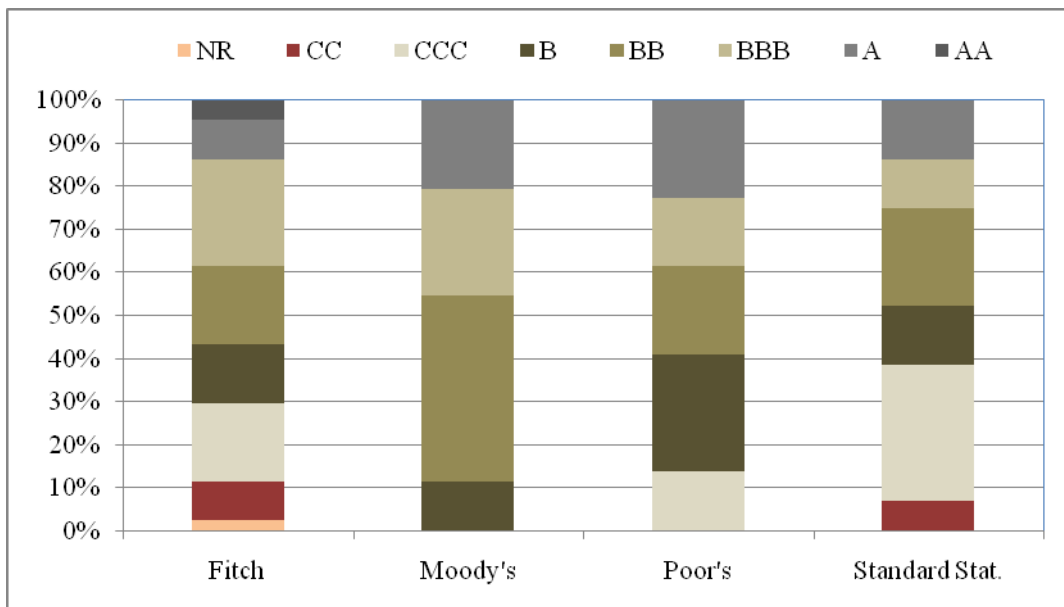
Note: an upgrade (resp. downgrade) measures any upward (resp. downward) revision occurring between previous year's manual and current year's manual. Publications dates are: January (Moody's), March: Poor's; August (Fitch) and December (Standard Statistics).

Figure 6
Cumulative accuracy profiles: 3 cases



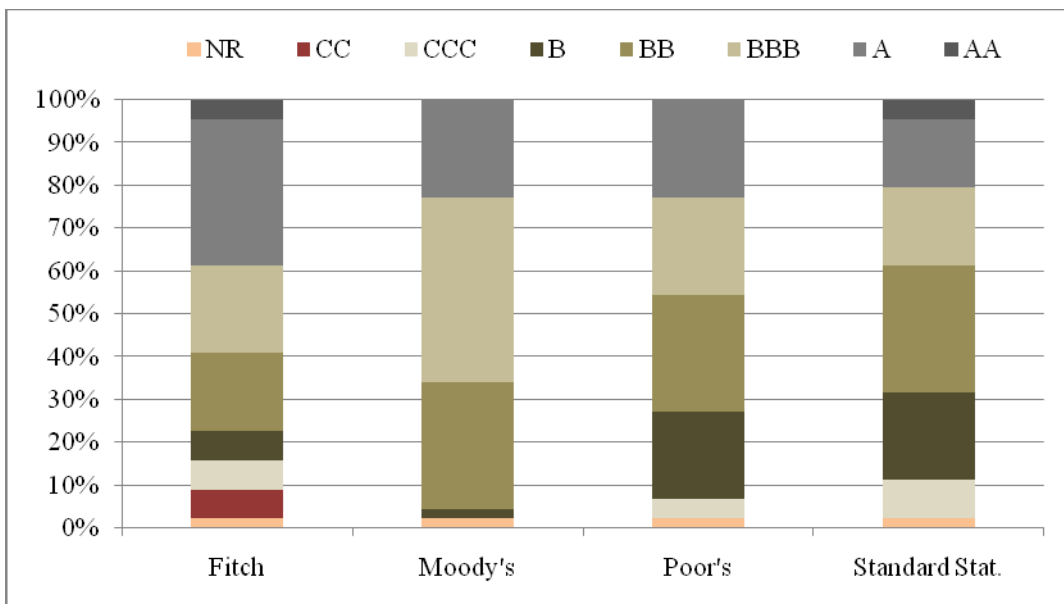
Source: See text.

Figure 7
Ratings composition at the time of default
 (1920–39, 44 defaulted issues)



Source: Authors' computations.

Figure 8
Ratings composition one year before the default
 (1920–39, 44 defaulted issues)



Source: Authors' computations.

Table 1

Products offered by the agencies

	Fitch	Moody's	Poor's	Standard Statistics
Name of the Main Publication Compiling Ratings	Fitch Stock and Bond Manual	Moody's Manuals of Investments	Poor's Volumes	Standard [Statistics] Bond Books
Frequency of the Main Publication	Annual	Annual	Annual	Monthly
Thickness of Annual Publications if Applicable	1930: 2,500 pages (a)	1930: 10,000 pages (b)	1930: 6,000 pages (c)	Not applicable
Price	\$35 (d)	\$100 (e)	Unknown	Not applicable
Publications with Other Frequencies	Daily, Semi-Weekly, Weekly, Monthly, Quarterly (f)	Daily, Weekly, Monthly	Four times a week, Weekly, Quarterly	Daily, Weekly, Monthly
Other Publications Starting	Daily, Semi-Weekly, Weekly, Monthly: 1928 (g) Quarterly: 1935 (h)	Weekly, Monthly: 1910 (i) Daily: 1932 (j)	Weekly: 1932 Four times a week: 1934 (k) Quarterly: 1934 (k)	Daily, Weekly, Monthly: 1922 (l)
Specific Publication Dedicated to Qualitative Analysis	Not until 1928 (m)	Yes (n)	Not until 1932 (o)	Yes (p)
Name and Frequency of the Specific Publication Dedicated to Qualitative Analysis	"Fitch Revisions", Semi-Weekly (q)	"Moody's Investment Letters", Weekly	"Poor's Analytical Services – Bonds & Stocks", Weekly	"Standard Daily Trade Service", Daily
Price of the Specific Publication Dedicated to Qualitative Analysis	\$150 (annual subscription) (r)	\$150 (annual subscription) (r)	\$120 (annual subscription) (s)	\$180 (annual subscription) (t)
Academic Advisers Members of the Board	No	No (u)	No	Yes (Profs. Scott, Davenport and Parry) (v)
Offices/Representatives Outside the United States (interwar)	Not until 1933 (w)	Yes (x)	Unknown	Yes (y)
Locations of Foreign Offices/Representatives if Applicable	Montreal and London (z)	London (x)	N.A.	London, Berlin, Paris, Geneva, Stockholm, Brussels, Panama Canal Zone, La Paz, Buenos Aires, Santiago, Warsaw (y)

Listed Company	No	Yes, from 1928	No	No
----------------	----	----------------	----	----

- (a) *Fitch Stock and Bond Manual 1930*.
- (b) Aggregation of *Moody's Railroad Manual 1930*, *Moody's Industrial Manual 1930*, *Moody's Public Utility Manual 1930* and *Moody's Government and Municipal Manual 1930*.
- (c) Estimation based on the consultation of various *Poor's Volumes* in 1929, 1930 and 1931.
- (d) *Fitch Stock and Bond Manual 1930* price.
- (e) Total sum of *Moody's Railroad Manual 1930*, *Moody's Industrial Manual 1930*, *Moody's Public Utility Manual 1930* and *Moody's Government and Municipal Manual 1930* prices.
- (f) *Fitch Bond Books'* introductions indicate "Daily, Semi-Weekly, Weekly, Monthly, Quarterly" publications, although we could not find these sources.
- (g) First indication found in *Fitch Bond Book 1928*.
- (h) First indication found in *Fitch Bond Book 1935*.
- (i) Babson (1910).
- (j) Based on the description of *Moody's Manuals* released in 1932.
- (k) First indication in *Poor's Industry Service – News, Facts, Forecasts* (1934).
- (l) First indication found in *Standard Daily Trade Service*, February 1922.
- (m) Based on the description of "Fitch Revisions" services found in *Fitch Bond Book 1928*.
- (n) Based on the consultation of *Moody's Manuals* and *Moody's Investment Letters* released from 1918.
- (o) Based on the consultation of *Poor's Volumes* published from 1922 and *Poor's Analytical Services – Bonds & Stocks* released from 1932.
- (p) Based on the consultation of *Standard Daily Trade Service* released from 1922.
- (q) Based on the description of "Fitch Revisions" services found in *Fitch Bond Book 1929*.
- (r) Subscription prices in 1929.
- (s) Subscription prices in 1923 (*Wall Street Journal*, 6 July 1923).
- (t) Subscription prices in 1931.
- (u) *Moody's Manuals* and *Investment Letters* published between 1922 and 1927 mention Max Winkler, PhD, but he is an employee of the agency.
- (v) See *Standard Daily Trade Service* publications released in 1922.
- (w) Found in *Fitch Bond Book 1933*.
- (x) Based on the consultation of *Moody's Manuals* and *Moody's Investment Letters* released from 1918.
- (y) Based on the consultation of *Standard Daily Trade Service* released from 1922.
- (z) Found in *Fitch Bond Book 1933* and *Fitch Bond Book 1938*.

Table 2
Ratings: data source

	Fitch	Moody's	Poor's	Standard Statistics
Business Starting	1913	1900	1868	1922
Ratings begin	1924	1909	1916	1922
Sovereign Ratings Begin	1924	1918	1922	1924
Name of Source	Fitch Bond Book	Moody's Manual of Investments – American and Foreign Government Securities	Poor's bank, government and municipal volume	Standard [Statistics] Bond Book
Frequency of Our Source	Annual	Annual	Annual	Monthly
Month of publication	August	January	March	December ^(a)
Separate Sovereign Volume (interwar)	No	Yes	Varies	No
Periods available (interwar)	1924-1939	1918-1939	1922-1939	1924-1939
Periods we used	1927-1936	1927-1936	1927-1936	1927-1936

^(a) Standard Statistics did not have an annual volume properly speaking. The choice of December from Standard Statistics monthly sources to compile annual data was arbitrary.

Source: Authors.

Table 3
Correspondence across ratings

Fitch	Moody's	Poor's	Standard Statistics	"Majority interpretation"	Moody's <i>Modern scale</i>
AAA	Aaa	A**	A1+	Highest	<i>Aaa</i>
AA	Aa	A*	A1	High	<i>Aa1, Aa2, Aa3</i>
A	A	A	A	Sound	<i>A1,A2, A3</i>
BBB	Baa	B**	B1+	Good	<i>Baa1, Baa2, Baa3</i>
BB	Ba	B*	B1	Fair	
B	B	B	B	Somewhat speculative	<i>Ba1, Ba2, Ba3</i>
CCC	Caa	C**	C1+	Speculative	<i>B1, B2, B3</i>
CC	Ca	C*	C1	Highly speculative	
C	C	C	C	Extremely speculative	<i>Caa1, Caa2, Caa3</i>
DDD	--	D**	D1+	Low or weak	<i>Ca</i>
DD	--	D*	D1	Small or very weak	<i>C</i>
D	--	D	D	Practically valueless	

Source: Harold (1938), p 75 and authors' computations.

Table 4

The determinants of Moody's ratings - regression results

Regressions	[1]	[2]	[3]	[4]	[5]
Date of ratings	September 1995	January 1925	January 1929	January 1925	January 1929
No. countries	49	37	43	37	43
Intercept	3.408 (1.379)	0.582 (0.350)	-0.788 (0.638)	-4.673 (0.792)	-5.626 (1.936)
Per capita income	1.027 (4.041)	1.926 (2.852)	2.339 (5.425)	3.347 (3.454)	5.026 (4.222)
Growth	0.130 (1.545)	0.083 (0.959)	0.031 (0.883)	0.630 (0.496)	0.037 (0.590)
Inflation	-0.630 (2.701)	-0.306 (2.534)	-0.044 (2.669)	-0.532 (2.993)	-0.097 (2.994)
Fiscal balance	0.049 (0.818)	-0.006 (0.808)	-0.242 (4.015)	-0.009 (0.841)	-0.422 (3.423)
External balance	0.006 (0.535)	-0.294 (3.412)	-0.327 (4.985)	-0.521 (3.495)	-0.665 (4.276)
External Debt	-0.015 (5.365)	-0.001 (3.851)	-0.002 (6.582)	-0.002 (2.462)	-0.005 (3.731)
Development indicator	2.957 (4.175)	0.545 (1.026)	0.130 (0.333)	0.789 (1.061)	0.673 (0.819)
Default history	-1.463 (2.097)	-0.852 (2.312)	-0.468 (1.680)	-1.158 (2.223)	-0.749 (1.485)
Adjusted R ²	0.905	0.762	0.841	/	/
Pseudo R ²	/	/	/	0.893	0.945

Notes: Absolute t-statistics are in parentheses. All results are significant at the 5% level. Coefficients with expected sign and significant t-statistic are in bold. Regression [1] is Cantor and Packer's results for 1995 Moody's ratings. Regressions [2] and [3] are the authors' OLS results. Regressions [4] and [5] are the authors' ordered probit results.

Table 5
1-year default rates (1928–37)

	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	Average 1928- 1937 ¹
Fitch											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	11	0	0	0	0	0	0	2
A	0	0	0	24	0	13	0	0	0	0	5
BBB	0	0	0	27	10	0	0	0	0	0	4
BB	0	0	0	0	56	0	13	0	25	0	9
B						33	0	0	0	25	10
CCC						80	33	0	50		56
IG	0	0	0	16	4	3	0	0	0	0	3
SG	0	0	0	0	56	23	10	0	25	13	16
Moody's											
AAA	0	0	0	0	0	0	0				0
AA	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	28	5	0	0	0	0	0	4
BBB	0	0	0	35	36	20	6	0	10	0	10
BB	0	0	0	17	60	45	20	0	0	17	19
B							0	0	50	0	17
CCC											
IG	0	0	0	16	7	8	4	0	4	0	4
SG	0	0	0	17	60	45	17	0	13	14	18
Poor's											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	18	4	8	0	0	0	0	4
BBB	0	0	0	30	4	0	0	0	5	0	5
BB	0	0	0	8	38	7	17	0	25	0	10
B	0	0	0	0	50	78	33	0	0	0	27
CCC							0	0	50	100	33
IG	0	0	0	17	3	4	0	0	2	0	3
SG	0	0	0	7	40	35	19	0	20	7	15
Standard Statistics											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	16	0	0	0	0	0	0	3
BBB	0	0	0	13	5	15	0	0	0	0	4
BB	0	0	0	38	5	0	33	0	10	11	7
B	0	0	0	29	30	0	0	0	33	0	13
CCC				50	100	73	33	0	50	0	59
IG	0	0	0	9	2	5	0	0	0	0	2
SG	0	0	0	38	22	23	23	0	20	8	16

¹ Weighted by the number of issues.

Note: For Tables 5–7, rating-based cohorts for each year are formed on the basis on the last observed rating the previous year. The reported rates for Tables 5–7 are the percentage of securities in each cohort that defaulted over that year (Table 5), that year and the next two years (Table 6), and that year and the next four years (Table 7).

Source: Authors' computations.

Table 6
3-year default rates
(1928–37)

	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	Average 1928-37 ¹
Fitch											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	6	16	6	0	0	0	0	0	5
A	0	17	28	45	25	13	0	0	0	0	18
BBB	0	26	64	60	30	0	0	0	0	0	18
BB	0	0	29	100	67	7	27	11	25	0	20
B						33	0	0	50	50	19
CCC						90	67	67	100		83
IG	0	13	22	31	19	3	0	0	0	0	13
SG	0	0	29	100	67	30	21	18	42	25	28
Moody's											
AAA	0	0	0	0	0	0	0				0
AA	0	0	0	0	8	0	0	0	0	0	1
A	0	25	28	45	32	0	0	0	0	0	17
BBB	0	33	75	82	73	28	9	4	10	0	28
BB	0	0	40	50	60	55	40	25	33	33	37
B							100	50	50	0	50
CCC											
IG	0	13	22	30	22	11	5	2	4	0	13
SG	0	0	40	50	60	55	50	30	38	29	38
Poor's											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0
A	0	15	18	24	21	8	0	0	0	0	11
BBB	0	24	40	52	19	0	0	0	5	0	17
BB	0	0	45	69	69	29	25	7	25	0	31
B	0	0	0	0	50	78	67	20	25	20	39
CCC							100	100	100	100	100
IG	0	14	20	26	15	4	0	0	2	0	11
SG	0	0	42	60	67	48	38	19	40	14	36
Standard Statistics											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	4	4	9	10	0	0	0	0	0	3
A	0	27	33	25	0	0	0	0	0	0	10
BBB	0	25	10	40	30	15	0	0	25	33	16
BB	0	18	64	50	11	5	50	13	20	11	22
B	0	0	71	86	80	33	25	33	33	0	51
CCC				83	100	82	67	100	50	0	78
IG	0	12	11	20	14	5	0	0	2	3	8
SG	0	14	67	71	41	31	46	27	27	8	37

¹ Weighted by the number of issues.

Source: Authors' computations.

Table 7
5-year default rates
(1928–37)

	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	Average 1928-37 ¹
Fitch											
AAA	0	0	0	4	0	0	0	0	0	0	1
AA	0	0	11	16	6	0	0	0	0	0	6
A	14	23	40	47	25	13	0	0	0	0	23
BBB	29	58	82	60	37	0	0	0	19	57	32
BB	67	50	43	100	78	15	27	11	25	0	28
B						33	0	0	50	75	24
CCC						100	100	100	100		100
IG	12	23	31	33	22	3	0	0	7	11	19
SG	67	50	43	100	78	38	24	23	42	38	37
Moody's											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	4	4	8	0	0	0	0	0	2
A	13	42	52	48	32	0	0	0	5	7	25
BBB	88	83	81	82	91	32	9	4	20	22	41
BB	0	0	40	50	80	73	60	38	33	33	46
B							100	50	50	0	50
CCC											
IG	15	25	32	32	25	13	5	2	10	14	20
SG	0	0	40	50	80	73	67	40	38	29	46
Poor's											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0
A	10	15	24	26	21	8	0	0	0	0	13
BBB	13	52	64	57	19	0	0	0	20	17	27
BB	71	50	55	69	85	36	25	7	25	38	46
B	0	0	0	0	100	100	100	40	25	20	55
CCC							100	100	100	100	100
IG	8	22	30	28	15	4	0	0	9	6	15
SG	63	44	50	60	87	61	44	24	40	36	50
Standard Statistics											
AAA	0	0	0	0	0	0	0	0	0	0	0
AA	5	4	8	9	10	0	0	0	0	0	5
A	13	47	61	25	0	0	0	0	0	0	16
BBB	13	33	30	53	30	15	0	0	25	100	23
BB	63	55	64	50	21	10	50	13	50	44	38
B	50	67	86	86	90	67	25	50	33	0	66
CCC				83	100	91	100	100	50	0	85
IG	7	17	21	23	14	5	0	0	2	9	12
SG	60	57	71	71	50	40	54	33	47	33	52

¹ Weighted by the number of issues.

Source: Authors' computations.

Table 8

The accuracy ratios for agency ratings and average rating

	Fitch	Moody's	Poor's	Standard	Average rating
1-year horizon					
1931	0.36	0.59*	0.32	0.56*	0.54*
1932	0.82	0.85	0.64	0.76	0.79
1933	0.81	0.73	0.67	0.73	0.74
1934	0.55	0.53	0.79	0.81*	0.87*
1936	0.90	0.65	0.75	0.91	0.77
Mean	0.69	0.67	0.64	0.75	0.74
Median	0.81	0.65	0.67	0.76	0.77
3-year horizon					
1929	0.46	0.62*	0.36	0.48	0.59*
1932	0.53	0.72*	0.51	0.58	0.63*
1935	0.74	0.74	0.85	0.89	0.95
Mean	0.58	0.70	0.57	0.65	0.71
Median	0.53	0.72	0.51	0.58	0.63
5-year horizon					
1929	0.65	0.76	0.63	0.68	0.75*
1934	0.66	0.75	0.81	0.82	0.88*
Mean	0.55	0.75	0.72	0.75	0.81

Note: The rating category for each year's securities is determined as in Tables 5–7. An accuracy ratio has an asterisk next to it if it is greater than at least one of the accuracy ratios of the other agencies (for the same yearly cohort) at the 95% level of statistical significance (test statistic from Hanley and McNeil (1983)).

Source: Authors' computations.

Table 9

Difference between the accuracy ratios for agency ratings and for yield-implied ratings

	Fitch	Moody's	Poor's	Standard	Average rating
1-year horizon					
1931	-0.22*	-0.10	-0.44*	-0.25*	-0.22*
1932	0.19	0.22	-0.08	0.00	0.06
1933	-0.06	-0.19*	-0.21	-0.18	-0.15
1934	0.00	0.01	0.13	0.20	0.30
1936	-0.04	-0.13	-0.18	-0.02	-0.18
Mean	-0.03	-0.04	-0.15	-0.05	-0.04
Median	-0.04	-0.10	-0.18	-0.02	-0.15
3-year horizon					
1929	0.10	0.33*	0.03	0.21	0.23
1932	-0.23*	-0.06	-0.31*	-0.22*	-0.18*
1935	-0.26	-0.13	-0.04	-0.04	-0.10
Mean	-0.13	0.05	-0.11	-0.02	-0.02
Median	-0.23	-0.06	-0.04	-0.04	-0.10
5-year horizon					
1929	0.05	0.25*	0.08	0.11	0.15
1934	-0.08	0.06	0.05	0.02	0.13
Mean	-0.01	0.16	0.08	0.08	0.14

Note: The rating category for each year's securities is determined as in Tables 4-6. The market yields upon which the yield-implied ratings were calculated is taken to be the yield of the security at the beginning of the year. A number has an asterisk next to it if it is significantly different from zero at the 95% level of statistical significance (test statistic from Hanley and McNeil, 1983), ie, the accuracy ratio of the rating agency was significantly higher (or lower) than that of the yield-implied rating.

Source: Authors' computations.

Appendix:

1. Example of Moody's ratings (Dominican Republic, 1927, p 444 of the Manual):

TITLE OF BOND	Interest Payable	Maturity	Amount Outstanding	Rating
1. Dominican Rep. Customs 5s of 1908.....	F&A 1	1958—S. F.	\$1,583,740*	A
2. Dominican Rep. Customs 5s of 1918.....	J&J 1	1938—S. F.	455,600	A
3. Dominican Rep. Customs 5½s of 1922.....	M&S 1	1942—S. F.	10,000,000	A

* As of Feb. 1, 1926.

2. Formula for the accuracy ratio (AR)

This formula is computed using the difference between the cumulative distribution of realisations and the cumulative distribution of ratings, or:

$$AR = 2 \left(\sum_{R_i=R_1, \dots, R_{\max}} \frac{(D_{R_i} + D_{R_{i-1}})(N_{R_i} - N_{R_{i-1}})}{2DN} - 0.5 \right)$$

where

D = total number of defaults

N = total number of issuers

D_{R_i} = total number of defaults rated R_i and less

N_{R_i} = total number of issuers rated R_i and less

$D_0, N_0 = 0$

$R_i = R_1, \dots, R_{\max}$ – ratings of a given agency

3. Formula for the yield-implied ratings

The goal is to compute the vector of brackets that minimise loss function $P(b)$. In the following formula, the $(x)^+$ operator takes value zero when the result within brackets is negative.

$$P(b) = \sum_j \left[w_j (s_j - b_{i(j)}^+)^+ + w_j (b_{i(j)}^- - s_j)^+ \right]$$

Where:

$i(j)$ = agency rating index of issue j .

s_j = spread of issue j over US bonds (for simplicity, since our computations are annual, and one average US yield is considered, we have performed computations over yields rather than spreads, without loss of generality).

$b_{i(j)}^+$ = upper threshold for rating j by agency i .

$b_{i(j)}^-$ = lower threshold for rating j by agency i .

w_j = rate chosen i j to level off the contribution of each rating bucket.