Reliability and Agreement of Credit Ratings in the Mexican Fixed Income Market

Ventura Charlin (*) and Arturo Cifuentes (§)

(*) V.C. Consultants

Los Leones 1300, Suite 1202

Santiago, CHILE

Telephones: (56) 22 789 7447 – (56) 97 989 8655

ventcusa@gmail.com

(§) CREM, Centro de Regulación y Estabilidad Macrofinanciera

Facultad de Economía y Negocios

Universidad de Chile

Santiago, CHILE

arturo.cifuentes@fen.uchile.cl

Working paper

NOVEMBER 2015

Abstract

Credit ratings play an important role in the fixed income market as the entire regulatory framework of this market segment is based on them and a significant part of what investors can and cannot do is dictated by ratings. Also, a number of ratings-based metrics are employed globally to estimate capital reserves, liquidity buffers, and solvency standards for many institutional investors such as insurance companies and pension funds. A critical assumption at the root of this regulatory architecture is that the credit-rating scales of the three leading agencies (Moody's, Fitch, and Standard & Poor's) are completely equivalent.

In this study we focus on the Mexican fixed income market. We find that the ratings of all three rating agencies exhibit a very high degree of inter-rater reliability. This means that in terms of ranking a group of bonds based on creditworthiness the three rating agencies would produce very similar results.

On the other hand, using a non-parametric statistic, the Wilcoxon matched-pairs test, we conclude that there are significant discrepancies among the ratings of the three agencies. This is consistent with a low level of inter-rater agreement detected. These findings challenge the suitability of credit ratings as a useful metric for regulatory purposes as they create the possibility of arbitrage.

KEYWORDS: credit ratings; rating agencies; bond markets; Mexican market; financial regulation; fixed income regulation.

JEL Codes: **G24** **G15** **F3**

1 Introduction

Credit ratings play a significant role in the fixed income markets. The main reason is that ratings remain entrenched in the regulatory framework of the U.S., the U.K., the Eurozone, and throughout Asia and Latin America. In fact, to a large extent, ratings dictate what many institutional investors such as insurance companies and pension funds can and cannot buy. Additionally, ratings influence many risk management-related metrics, determine which assets can be used as collateral in many two-party transactions, and affect the liquidity and price of the securities that are downgraded. Moreover, less sophisticated investors still rely on ratings as a guiding benchmark to make investment decisions. Also, credit ratings' downgrades, especially when they reach the below-investment grade area, can force investors to sell their holdings under unfavorable market conditions. Accordingly, credit ratings have an extraordinary power to influence the dynamics of all segments of the bond markets: corporate, sovereign, and structured products.

The credit rating agencies −notwithstanding the setback to their reputation they suffered after the subprime crisis− still enjoy a strong oligopoly power. Moody's, Standard & Poor's (S&P), and FitchRatings (Fitch) command in aggregate a market share that exceeds 95% of the global ratings market (Matthies 2013). There are probably many reasons behind this situation. One of the reasons is that regulators in both, the U.S. and Europe –perhaps afraid of creating more havoc in the aftermath of the subprime crisis– decided not to pursue any serious actions against the rating agencies after the subprime crisis. This is in clear contrast, for example, with the Enron scandal which resulted in Arthur Andersen surrendering its license and ceasing to exist. Another reason is that regulators have maintained huge barriers to entry which obviously have protected the established players. Consider the case of the U.S. Securities and Exchange Commission (SEC), which controls access to the American market by virtue of designating which organizations can qualify as Nationally Recognized Statistical Rating Organizations (NRSRO). To obtain the desirable NRSRO status, a firm must prove −among other things− that it has been in the business of issuing paid ratings and selling them to an established group of clients for at least three years before being able to apply for such status. This requisite presents a difficult obstacle to overcome for a potential newcomer for it must find a market for a product (a rating) that has yet to be officially sanctioned.

With this background, it is not surprising that credit ratings have received an increasing amount of attention from the academic environment. Previous research has addressed a number of issues. Given that credit ratings are supposed to reflect the degree of creditworthiness of an issuer, many studies have investigated if ratings indeed capture this feature and results have shown that over several-year periods credit ratings correlate well with observed default rates. However, in shorter (one-year) periods, ratings are a less reliable predictor of default (Zhou 2001; Jorion and Zhang 2007).

Several authors have explored the capacity of credit ratings to influence market movements. They have concluded that ratings decisions, especially downgrades, influence prices and liquidity. They have speculated that the main reason is that fixed income markets are not very efficient and credit ratings, despite their shortcomings, convey some information that is not readily available to all investors (Gonzales et al. 2004). Fabozzi and Vink (2015), in the context of European mortgages, established that investors place some value on ratings as they penalize transactions in which the three agencies give different ratings to the same tranches.

Other researchers have attempted to identify the factors that determine credit ratings. In principle, credit ratings should capture quantitative factors such as financial ratios, combined with qualitative factors such as macroeconomic drivers and corporate governance elements. The relationship between certain metrics (long-term debt, interest coverage ratios, as well as their evolution over time) and creditworthiness has been well-established. Nevertheless, how to properly combine these quantitative metrics with qualitative factors to estimate credit ratings is still an area with many unresolved issues (Altman 1968; Carling et al. 2007; Kamstra et al. 2001).

Scholars have examined the stability of ratings over time and through the business cycle, as well as the transition-probability matrices, that is, the likelihood that a given rating could change to a different category during a specific time-frame, typically, one year (Amato and Furfine 2004; Kim and Sohn 2008; Loffler 2004, 2005). It should be noted that most research has focused on corporate ratings, and less emphasis has been given to sovereign ratings and structured products.

Matthies (2013) has provided the best survey of previous credit ratings-related empirical research. However, what is really surprising is that except for a recent report by Ghosh (2013) nobody has studied the degree of agreement (or disagreement) among the corporate ratings provided by the three leading agencies. In other words, the assumption that the rating categories used by the three leading agencies are equivalent has been accepted without question by regulators, politicians, and naïve investors.

At this point, it is necessary to introduce a few facts regarding the ratings scales and their meaning:

[1] S&P and Fitch give ratings or opinions regarding the creditworthiness of an issuer based on a default probability (P) criterion, i.e., the likelihood that the issuer might not be able to pay its future obligations. Moody's claims that its ratings are based on the concept of expected loss (EL); that is, they take into account both, the default probability and the loss given default (LGD) or, alternatively, the recovery rate or market value of the defaulted security. Based on these considerations, one should expect some degree of

divergence between the ratings given by Moody's and the other two agencies, especially for securities associated with very low or very high recoveries.

[2] Fitch and S&P use the same symbols to designate their rating categories (AAA, AA, A, etc.), whereas Moody's uses a slightly different notation (Aaa, Aa, A, etc.). For the purpose of this study we will consider seventeen rating categories. Table 1 shows the ratings scales in detail.

[3] These scales are ordinal as their categories are not equally spaced. Although the agencies have stated clearly that their scales reflect decreasing levels of creditworthiness, they have been reluctant to specify precise cutoff points (either in terms of P or EL) between the rating categories. The empirical evidence (observed default rates by categories) indicates that the overall trend is consistent: lower ratings correspond to higher default rates (see Table 1). However, there are some anomalies, for instance, securities rated AA- by Fitch have performed better than those rated AA, and Baa1-rated bonds by Moody's have experienced lower defaults that those with a A2 rating. Also, a visual inspection of the observed default rates by category seems to indicate that the existence of a one-to-one correspondence among categories is not apparent.

[4] Regulators, and to a less extent investors, implicitly assume that there is a solid one-to-one equivalence between the rating categories of the different agencies. For example, the definition of investment grade (essentially, the level above which institutional investors are allowed to buy securities) means either having a BBB- rating or a Baa3 rating. In other words, the assumption is that a BBB- by S&P, a BBB- by Fitch, and a Baa3 by Moody's, have all the same meaning. Another case in point, typically, bonds with ratings of either Aa3 or AA- (by either S&P or Fitch) are assigned always the same haircut value in tables dealing with fixed income securities held as collateral.

For all practical purposes, it has been assumed that the three credit ratings scales are equivalent in the sense that they convey the same amount and type of information. Consequently, a violation of this assumption would have far reaching implications for it would weaken the conceptual foundation of much of the regulation of an important segment of the capital markets.

To our knowledge, the only study that has challenged the notion of ratings equivalence is a recent report by Ghosh (2013). The author investigated the differences between corporate ratings issued by Moody's and S&P in the U.S. market. He considered the companies in the Russell 3000 index that had been rated by both, S&P and Moody's. The study considered six specific dates in the 2006-2012 timeframe and concluded that Moody's ratings were consistently lower that S&P's. Somehow contrary to expectations, the author found that in industries that normally enjoy high and low recovery rates, the ratings differences were minimal.

The focus of our study is the equivalence among credit ratings in the Mexican fixed income market by all three agencies. Leaving aside that no such study has been conducted before, the Mexican market presents several interesting features. First, it is the most developed fixed income market in Latin America. Second, Mexico has carried out many market-liberalization reforms aimed at establishing a free-market economy, albeit with some setbacks. And third, its economy (2013 GDP = US$ 2 trillion at PPP) is poised to soon overtake Brazil as the largest in Latin America. Finally, Mexico, which has been an investment grade country since 2002 and a member of the OECD since 1994 (the first Latin American country to join this exclusive organization) is a good emerging market success story. All these considerations make it relevant to explore the ratings-equivalence conundrum in this market.

2 The Mexican fixed income market[1]

From 1994 until 2012, the period considered in this study, Mexico grew both, in terms of population and economic output. Its population increased from 89.6 million in 1994 to 114.9 million in 2012; and its GDP (at PPP) increased from US$ 785 billion to US$ 1,758 billion in the same period. And the market capitalization of listed companies in 2012 was equivalent to 44.3% of its GDP (in 1994 it was 24.7%).

In 2013, Mexico's total debt issuance (corporate plus sovereign), accounted for 34% of total Latin American debt issuance (Brazil was second with 30%). More interesting, Mexico's bond market has always enjoyed a high level of foreign participation: as of the end of 2013 the figure was around 40%; this is in contrast to Brazil, the other big Latin American market, where foreign ownership of local bonds has never surpassed the 20% level. In 2003, in relation with its government bonds contracts, Mexico pioneered a set of then-innovative collective action clauses that were adopted by many emerging markets nations. At the end of 2014, the volume of exchange rate derivatives in the Mexican market reached US$ 25 billion, from an almost zero level in 2005.

[1] Information regarding Mexico gathered from the following sources:

Central Intelligence Agency. (2014). The World Factbook 2013-2014, Washington D.C.
Banco de Mexico. (2014). Compilacion de Informes Trimestrales Correspondientes al año 2013, available from http://www.banxico.org.mx/publicaciones-y-discursos
Kemen, G. (2013). Emerging Debt Markets: Still On Strong Footing, Americas Quarterly, July.
Moore, E. (2014). Mexico's Move Set to Shake Up Bond Market, Financial Times, November 11.
World Bank. (2014). Information available from http://data.worldbank.org/country/mexico

In 2000, Moody's granted Mexico investment grade status, the first rating agency to do so; Fitch and S&P followed suit in 2002. By 2012, the country sovereign ratings were BBB/Baa1/BBB according to Fitch, Moody's, and S&P respectively. In 1994 the sovereign country ratings were BB/Ba2/BB+.

The number of corporate bond issuances in the period 1994-2012 (counting based on CUSIP numbers obtained from the Bloomberg database) was 1,015. In terms of 2010 US dollars, these issues amount to a total of US$ 297.3 billion. Of these 1,015 issuances, 861 (85%) were rated by at least one of the three leading agencies. In terms of dollar volume, the percentage rated at issuance was 97.2%.

Of the 1,015 issuances, 376 were in local currency (37%); whereas in volume, the local currency issuance was only 2.1% (slightly more than US$ 6 billion). Foreign currency bond issuances were dominated by US dollars (93%), followed by Euros (4%).

Fitch, Moody's and S&P rated, respectively, 455, 731 and 551 corporate bond issuances in this period. In terms of 2010 US dollars the corresponding figures were US$ 236 billion, US$ 259 billion and US$ 253 billion. Table 2 summarizes the basic information regarding bond issuances in the 1994−2012 period. Figures 1 and 2 show additional information regarding the rated debt market.

It can be seen that no agency dominates any particular market segment, except for Moody's that enjoys a much higher fraction of the financial sector. Also, the number of new issuances has grown more or less steadily since 2002 (the year Mexico became investment grade according to all three agencies) and the subprime crisis (2008) made only a small dent in the issuance volume.

In the Mexican bond market, as is typically the case in most fixed income markets, credit ratings play an important role in the regulatory architecture of this market segment[2]. For instance, the capital requirement rules stated by the Comisión Nacional Bancaria y de Valores (the equivalent of the SEC in the U.S.) which affect banks, savings and loans institutions, and economic development institutions, are all based (when it comes to credit-related instruments) on credit ratings. Moreover, the numerous tables imbedded in the text of the relevant documents make explicit the ratings-equivalence assumption already mentioned. A similar situation occurs in the regulatory norms stated by the Comisión Nacional de Seguros y Fianzas (the insurance regulator) in relation to admissible investments and counterparty risks. Also, the investment guidelines for the AFORES (the institutions that manage the pension funds) are structured around limits based on credit ratings. (The AFORES, whose assets under management are equivalent to almost 15% of

[2] The information referred to in this paragraph has been gathered from the Central Bank of Mexico website (www.banxico.org/mx), the Secretaría de Hacienda y Crédito Publico website (www.hacienda.gob.mx), and the September 19, 2015 edition of El Economista (www.eleconomista.com.mx).

the Mexican GDP, have more than 90% of their funds invested in rated debt instruments.) Finally, with Mexico making important advances to adopt Basel III and Solvency II standards, which rely both on credit ratings for an important part of their solvency and liquidity rules, the relevance of exploring the validity of the ratings-equivalence hypothesis is paramount.

3 The data

Our data are based on corporate ratings, at issuance, as reported by Bloomberg. We focus on the 1994-2012 period. The advantage of restricting the study to at-issuance ratings (as opposed to analyzing all the ratings at some specific time point) is that it makes for a clean comparison. In fact, it eliminates the possibility that any observed discrepancy could be attributed –at least potentially– to different standards of attention in terms of monitoring existing ratings. Or the possibility that one rating agency (a stricter agency) might lead in terms of downgrading an issuer, which, in turn, might encourage the other agencies to alter their ratings not to appear as more benevolent. In other words, focusing on ratings released simultaneously and with access to the same information, assures us that whatever discrepancy is detected it will be only the result of applying different criteria to the same evidence.

Table 3 describes the data from the three rating agencies: Moody's, S&P, and Fitch. We have collapsed all the observations with the CCC, CC, C and D (S&P and Fitch nomenclature) designation into one category simply because the agencies tend to report the data for these categories in aggregate terms. Rating agencies make little or almost no difference between, for example, a CCC-rated bond and a C-rated bond. Additionally, although strictly speaking these scales are ordinal, we have assigned a numeric rank of 1, 2, etc. to each category to facilitate comparisons.

4 Analysis and discussion

Let us first emphasize that the rating scales are just ordered categories, and therefore, an average ranking is not a meaningful concept. Notwithstanding this caveat, a warning is appropriate: a naïve inspection of Table 3 might be interpreted as an indication that the three agencies give similar ratings and therefore, their scales are equivalent. After all, 10.23, 10.82 and 11.44 (the average numerical rankings) do not seem that different. This conclusion, as we will see shortly, could be misleading.

Before we continue, we need to specify what we mean when we talk about the ratings being equivalent since this is not a term of art. In principle, there are several criteria (tests) that can be used to address this question, and each criterion illuminates a different aspect of what "equivalent" could mean. In this study, we employ three such criteria: (i) inter-rater reliability; (ii) inter-rater agreement; and (iii) differences of rankings in paired-observations.

In this context, each rating agency is considered to be an independent judge. This assumption is probably reasonable as we are dealing with three ratings that are issued simultaneously and with no interchange of information among the judges. On the other hand, the judges (rating agencies) are not single individuals, but a group of individuals (rating committees) whose composition obviously changes over time. This effect, however, is mitigated by the fact that these individuals follow certain established ratings criteria which give stability and continuity to the rating process.

4.1 Inter-rater reliability

This concept (which is often confused with inter-rater agreement) refers to the degree to which ratings produced by different judges are identical when expressed as deviations from their means. Or, alternatively, it means the degree to which the order relationship implied by the ratings of one judge is analogous to the order relationship implied by the ratings of a second judge (regardless of the numerical value assigned to the ratings).

The inter-rater reliability can be examined using the R_{SF} coefficient (Shrout and Fleiss, 1979) which is defined as

$$R_{SF} = \frac{MS_{IS} + MS_e}{MS_{IS} + MS_e(K-1)}$$

where the analysis involves using the standard two-way ANOVA procedure given the assumption of no interaction between issuances and judges (rating agencies), to compute the mean square for issuances (MS_{IS}) and the mean square for error (MS_e). These components are then inserted into the standard equation for reliability, and K = the number of agencies rating each issuance.

Values closer to 1 indicate a high degree of reliability whereas values closer to 0 show the opposite. Table 4 displays the R_{SF} values for all three possible comparisons. Clearly, we have a high degree of reliability. This means that the ordinal relationships implied by the ratings of the three agencies are very similar. In other words, if we just want to know how three bonds –X, Y and Z– are ranked, based on their creditworthiness, we should be indifferent in terms of which rating agency must be used.

4.2 Inter-rater agreement

This concept refers to the degree to which two judges or agencies, tend to assign the same ratings, to each one of the issuers considered. Thus, inter-rater agreement captures also the differences between the ratings assigned. Inter-rater agreement can be tested by means of the T_{TW}-index coefficient (Tinsley and Weiss, 1975) which is defined as

$$T_{TW} = \frac{N_a - N \times \rho_c}{N - N \times \rho_c}$$

where N_a = the number of agreements, N = the number of issuances rated, and ρ_c = the probability of chance agreement on an issuance. Positive (negative) values are associated with levels of agreement which are higher (lower) than the agreement one could have obtained simply by chance. Table 5 shows the T-index values for three comparisons among the rating agencies.

The first and second panel define agreement in reference to the 17-rating categories specified in Table 1. Agreement, in the first panel, means having exactly the same rating (0 points of discrepancy). In the second panel, we consider a slightly more relaxed version of agreement, 0 or 1 points of discrepancy, except for ratings categories 10 and 11 (BBB- and BB+) where a 0 point agreement is required. The reason is that the BBB-/BB+ boundary marks the difference between investment grade and speculative bonds: a critical distinction for bond investors and regulators.

The third panel defines agreement as a 0 point difference but in reference to the broader 7-rating categories (AAA, AA, A, BBB, etc.) that is, we collapse, for example, the BBB+, BBB and BBB- rated bonds in one category; the same for the AA+, AA and AA-, and so on.

The T_{TW}-values suggest a degree of inter-rater agreement which is higher than chance, but still low-to-moderate. As expected, the second and third panels indicate higher inter-rater agreements which are obviously consistent with the more relaxed definitions of agreement employed. An interesting feature is that the agreement between S&P and Fitch is always higher than the agreement between Moody's and either S&P or Fitch. This situation is more salient (and more telling) in the first panel. This can be attributed to the fact that Moody's bases its ratings on the EL concept whereas its competitors rely on the P concept. In essence, Moody's employs a different benchmark than its competitors to estimate ratings.

A direct consequence of this low-to-moderate agreement is that if we were interested in inferring –for example– what is the default probability of a given bond, based on its rating, we would arrive at a fairly different conclusion based on the rating agency employed. See Table 1. Based on historic performance, a bond rated Aa3-by Moody's might suggest that its 10-year default rate could be 1.08%; however a Fitch or S&P AA- rating, might suggest different estimates: 0.21 and 0.79% respectively. This type of inter-rater disagreement is critical for it creates the opportunity for regulatory arbitrage, i.e., ratings from different agencies do not have identical numerical attributes.

4.3 Ranking differences in paired-observations

The idea here is to investigate whether the median difference between paired-observations (ratings) is zero (null hypothesis) using the Wilcoxon matched-pairs signed-ranks test (Wilcoxon, 1945). The Wilcoxon signed-rank test is the nonparametric alternative method to the paired sample t-test; paired observations are presumed to be drawn at random from a single population and the distribution of a random variable $D = X_1 - X_2$ has median zero. The Wilcoxon signed-rank test makes the additional assumption that the distribution of D is distributed symmetrically about the median. Tables 6.1, 6.2, and 6.3 describe the three possible rating agency comparisons. The Wilcoxon matched-pairs signed-ranks test the hypothesis that two distributions are the same, that is, $X_1 \sim X_2$.

Let d_j denote the difference for any matched pair of observations, $d_j = x_{1j} - x_{2j}$ for $j = 1, 2, \ldots, n$; and let N_r be the reduced sample size after excluding the pairs where $|x_{1j} - x_{2j}| = 0$. The Wilcoxon matched-pairs signed-ranks test statistic is defined as

$$W = \sum_{j=1}^{N_r} R_j$$

where $R_j = \text{sgn}(d_j)\ \text{rank}(|d_j|)$.

We analyze the aggregate data in each case and also perform comparisons by industry sectors as well as issuance-size in U.S. dollars. Notice that since the rating scales are ordinal (ranked data) −not interval, i.e., with equally spaced ordered categories− a comparison using a parametric paired differences t-test would be inappropriate.

Tables 6.1 and 6.3 show (see bottom right corner) that overall Moody's and S&P ratings as well as S&P and Fitch ratings are not equivalent (they show disagreement). The comparison between Moody's and Fitch (Table 6.2), at the aggregate level, shows no significant difference between the two ratings scales distributions. However, looking in more detail, we appreciate that the discrepancies between the ratings distributions are more noticeable when associated with large-size issuances. In fact, in all three comparisons, we notice that for issuances larger than US$ 200 million (which amount for more than 60% of the total volume) we see significant discrepancies (see the corresponding signed-rank score). Also, in all three cases, in the smallest issuance-size segment (less than US$ 20 million), the differences are not significant. This situation might be the result of having a small number observations in this segment.

In principle, we could have expected a high level of agreement between S&P and Fitch, since both give ratings based on the same concept: default probability. Table 6.3, as already mentioned, indicates that this

is not the case. By the same token, since Moody's gives ratings based on an expected loss concept, we could have anticipated a substantial difference between Moody's and the other two agencies. This is the case between Moody's and S&P (Table 6.1, S=−2955, p<.05). However, between Moody's and Fitch, at the aggregate level, the hypothesis of no difference in the rankings' distribution could not be rejected. On the other hand, if we look at the industries in which Moody's and S&P, and, Moody's and Fitch, show significant differences (communications, energy, and industrial & materials) we detect that the corresponding gaps coincide in their signs. This might suggest that when there are rating discrepancies between Moody's and its two competitors, these discrepancies could be attributed to the different rating approaches, namely, default probability (S&P and Fitch) versus expected loss (Moody's).

We also considered those instances where ratings by all three rating agencies were available. Follow-up pairwise comparisons were conducted using, again, the Wilcoxon signed-rank tests as indicated in Table 7. The results are consistent with those reported in Tables 6.1, 6.2 and 6.3. We also conducted a non-parametric Friedman test of differences among the rankings given by the three rating agencies. The differences among the rankings are statistically significant. The test is significant $\chi^2(2, N = 308) = 50.21$, $p < .0001$ for the ratings at issuance.

Considering that Mexico became an investment-grade country in 2002, it makes sense to look at the dataset by dividing it in two groups: (i) before, and (ii) after such event. Table 8 shows the results of the three pairwise comparisons, for both time periods. In the 2003−2012 period all three comparisons are significant. Not only that, they show the same tendencies (gap signs) as the results displayed in Tables 6.1, 6.2 and 6.3. In the period 1994−2002 the differences are also significant. There is, however, a noticeable fact. While S&P and Fitch maintained the same type of discrepancy, before and after 2002, (namely, S&P gave lower ratings) the opposite happens between Moody's and its two competitors. More precisely, initially Moody's tended to give lower ratings than both Fitch and S&P, but apparently during the second period, there was a ratings' standards changed and the opposite is observed. One could have suspected that the Moody's-Fitch discrepancy was weak based on the results reported in Table 6.2. But, as the results in Table 8 indicate, this is not the case. They suggest that Moody's possibly changed standards after 2002, and went from more severe to more lenient (note the change in sign in the gap). Thus, one can speculate that the difference between Moody's and its competitors might appear mitigated when analyzing the entire dataset (1994-2012) simply because the discrepancies before and after 2002 might be "cancelling out."

In summary, although the ratings by the three agencies exhibit a high level of reliability, the inter-rater agreement and the Wilcoxon test indicate that the three rating agencies give ratings that cannot be

considered to be equivalent as their respective distributions are dissimilar. Furthermore, our analyses suggest that S&P is consistently stricter than both Moody's and Fitch.

5 Conclusions

Two conclusions emerge from this study. First, the bad news: the lack of inter-rater agreement, coupled with the discrepancies detected by the Wilcoxon matched-pair test, raises some issues regarding the suitability of ratings for regulatory purposes. Notice that the entire fixed income regulatory framework in Mexico is based on the ratings-equivalence assumption. For example, rules such as, "cannot invest in any assets with a rating below Baa3/BBB-", or, "can only hold AAA/Aaa assets" and the like, are at the core of the country's current regulation. Violation of the ratings-equivalence assumption undermines the validity of such statements. Moreover, it encourages regulatory arbitrage, and introduces distortions that might affect the risk metrics employed to assess the solvency and liquidity of key financial institutions.

The good news is the high inter-rater reliability. This means that any ratings scale (Fitch, S&P or Moody's) is essentially identical if we just need to rank a group of bonds by creditworthiness. This is useful for an investor who needs to decide between any two bonds based on credit risk only.

The lack of inter-rater agreement to some extent should not be surprising. Not only S&P and Fitch use a criterion (P) that is different than that employed by Moody's (EL); but also, the consistency over time of rating committees (group of individuals whose composition vary) might increase the likelihood of disagreement.

One potential solution to this situation is to rethink the fundamentals of the regulatory framework and consider the possibility that the regulator might specify a well-defined ratings scale, independent of the agencies themselves. For example, a 17-category scale specified by clearly defined cutoff points based on different P (or EL) levels. In this case, the agencies should only limit themselves to issue an opinion regarding which of the categories should be assigned to a new bond issuance. This topic is clearly beyond the scope of the present study.

In case these findings turned out to be a feature not only of the Mexican market, but a general feature of the global fixed income markets, the consequences could be more unsettling. There is reason to believe that such could be the case. The recent study by Ghosh (2013) regarding the U.S. market, although it only considered ratings of companies listed in the Russell 3000 index, is revealing. This author examined only S&P and Moody's ratings and concluded that there were significant discrepancies between both agencies. At present, we are investigating if our findings regarding the Mexican market hold true in the broad U.S. fixed income market.

REFERENCES

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance*, 23(4), 589−609.

Amato, J., & Furfine, C. (2004). Are credit ratings procyclical?, *Journal of Banking & Finance*, 28, 2641−677.

Carling, K., Jacobson, C., Linde, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macroeconomy, *Journal of Banking & Finance*, 31(3), 845−868.

Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1), 37–46.

Fabozzi, F.J., & Vink, D. (2015). The information content of three credit ratings: the case of European residential mortgage-backed securities, *The European Journal of Finance*, 21 (3), 172−194.

Ghosh, S. (2013). A study of differences in Standard & Poor's and Moody's corporate credit ratings, unpublished report, Leonard Stern School of Business, New York University.

Gonzales, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Zins, C., & Wieland, M. (2004). Market dynamics associated with credit ratings: a literature review, *Banque de France Financial Stability Review*, 4, 53−76.

Jorion, P., & Zhang, G. (2007). Information effects of bond rating changes: the role of the rating prior to the announcement, *The Journal of Fixed Income*, 16(4), 45−59.

Kamstra, M., Kennedy, P., & Suan, T−K. (2001). Combining bond rating forecasts using logit, *The Financial Review*, 37, 75−96.

Kim, Y., & Sohn, S. (2008). Random effects model for credit rating transitions, *European Journal of Operational Research*, 184, 561−573.

Loffler, G. (2004). An anatomy of rating through the cycle, *Journal of Banking & Finance*, 28, 695−720.

Loffler, G. (2005). Avoiding the rating bounce: why rating agencies are slow to react to new information, *Journal of Economic Behaviour & Organization*, 56, 365−381.

Matthies, A. (2013). Empirical research on corporate credit-ratings: a literature review, Humbolt-Universitat zu Berlin, SBF 649 Discussion Paper 2013-003.

Shrout, P.E., & Fleiss, J.L (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

Tinsley, H. E. A. & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgements. *Journal of Counseling Psychology*, 22, 358-376.

Tinsley, H. E. A. & Weiss, D. J. (2000). Interrater reliability and agreement. In Tinsley, H. E. A. & Brown, S. D. (Eds.). *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 94-124). New York: Academic Press.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods, *Biometrics Bulletin*, 1, 80−83.

Zhou, C. (2001). Credit Ratings and Corporate Default, *The Journal of Fixed Income*, 3(11), 30−40.

Table 1 Rating Categories and 10-Year Cumulative Default Rates for: Moody's, Fitch, and S&P

Moody's Rating Categories	Fitch and S&P Rating Categories	Moody's Cumulative 10-Year Default Rates (in percent)[1]	Fitch Cumulative 10-Year Default Rates (in percent)[2]	S&P Cumulative 10-Year Default Rates (in percent)[3]
1. Aaa	1. AAA	0.18	0.00	0.74
2. Aa1	2. AA+	0.25	0.00	0.50
3. Aa2	3. AA	1.30	0.36	0.99
4. Aa3	4. AA−	1.08	0.21	0.79
5. A1	5. A+	2.33	0.89	1.29
6. A2	6. A	3.42	2.05	1.69
7. A3	7. A−	3.27	2.53	1.74
8. Baa1	8. BBB+	2.66	2.39	2.73
9. Baa2	9. BBB	4.50	4.79	3.91
10. Baa3	10. BBB−	6.87	7.54	6.84
11. Ba1	11. BB+	13.87	10.15	9.05
12. Ba2	12. BB	14.01	13.78	13.39
13. Ba3	13. BB−	29.49	9.19	18.33
14. B1	14. B+	36.85	10.12	24.25
15. B2	15. B	39.96	13.97	27.67
16. B3	16. B−	46.84	10.19	32.94
17. Caa1/Caa2/ Caa3/Ca/C	17. CCC+/CCC/ CCC−/CC/C/D	79.35	39.54	51.35

[1]Source: Moody's Investors Service Average Cumulative Issuer-Weighted Global Default Rates by Alphanumeric Rating, 1983-2013 from "Corporate Default and Recovery Rates, 1920-2013" (February 2014).

[2]Source: Fitch Global Corporate Finance Average Cumulative Default Rates: 1990–2013 from "Fitch Ratings Global Corporate Finance 2013 Transition and Default Study" (March 2014).

[3]Source: Standard & Poor's Global Corporate Average Cumulative Default Rates By Rating Modifier (1981-2013) from "Default, Transition, and Recovery: 2013 Annual Global Corporate Default Study and Rating Transitions".

Table 2 Corporate Debt Rated at Issuance by at Least One Rating Agency (RA)

		Volume in 2010 US$ Billion	Percentage of Total
Volume:			
	TOTAL	$288.97	
	In Local Currency	$5.39	1.9
	In Foreign Currency	$283.58	98.1
		Number of Issuances	**Percentage of Total**
Issuances:			
	TOTAL	861	
	In Local Currency	267	31.0
	In Foreign Currency	594	69.0
	Investment Grade [by at least one RA]	480	55.7
Industrial Sector:			
	Communications	197	22.9
	Consumer	136	15.8
	Energy	113	13.1
	Financial	222	25.8
	Industrials & Materials	149	17.3
	Other	44	5.1
Issue Size:			
	< US$ 20 Million	193	22.4
	US$ 20 - US$ 200 Million	219	25.4
	US$ 200 - US$ 400 Million	219	25.4
	US$ 400 plus Million	230	26.7
Rated by:			
	Moody's	731	84.9
	Fitch	455	52.8
	S&P	551	64.0
Rated by Two RA:			
	Moody's and S&P	455	52.8
	Fitch and Moody's	354	41.1
	Fitch and S&P	375	43.6
Rated by All Three RA:		308	35.8

Table 3 Number of Issues Rated at Issuance by: Moody's, Fitch, and S&P

Moody's Rating Categories	Fitch and S&P Rating Categories	Numerical Rank	Number of Issues Rated At Issuance by: Moody's	Fitch	S&P
1. Aaa	1. AAA	1	15	4	13
2. Aa1	2. AA+	2	0	0	0
3. Aa2	3. AA	3	3	0	1
4. Aa3	4. AA−	4	4	0	0
5. A1	5. A+	5	6	0	0
6. A2	6. A	6	60	17	0
7. A3	7. A−	7	63	33	39
8. Baa1	8. BBB+	8	168	33	32
9. Baa2	9. BBB	9	35	86	56
10. Baa3	10. BBB−	10	41	66	72
11. Ba1	11. BB+	11	44	18	30
12. Ba2	12. BB	12	67	51	87
13. Ba3	13. BB−	13	76	56	66
14. B1	14. B+	14	39	57	72
15. B2	15. B	15	56	7	47
16. B3	16. B−	16	39	11	24
17. Caa1/Caa2/ Caa3/Ca/C	17. CCC+/CCC/ CCC−/CC/C/D	17	15	16	12
		TOTAL	731	455	551
		Average Numerical Rank*	10.23	10.82	11.44
		Std. Deviation of Rank*	3.51	2.89	3.09

*Rating categories were assigned a numerical rank from 1 to 17 in ascending order; the rating category AAA/Aaa was assigned the lowest ranking and the rating categories CCC+/Caa1 through D/C were assigned the highest ranking.

Table 4 Inter-Rater Reliability: Comparisons of Ratings

		Ratings		
		Moody's vs. S&P	Moody's vs. Fitch	S&P vs. Fitch
Inter-Rater Reliability	R_{SF}	0.918	0.869	0.934

Table 5 Inter-Rater Agreement: Comparisons of Ratings

		Ratings		
		Moody's vs. S&P	Moody's vs. Fitch	S&P vs. Fitch
T_{TW}-INDEX Agreement (0 point discrepancy)	T-Index	0.302	0.195	0.520
	$\chi 2$ Value	495.2	159.4	1211.7
	P-Value	0.000	0.000	0.000
T_{TW}-INDEX Agreement (1 point discrepancy)	T-Index	0.682	0.656	0.812
	$\chi 2$ Value	753.0	542.2	880.1
	P-Value	0.000	0.000	0.000
T_{TW}-INDEX Agreement (0 point discrepancy) for Broader Categories	T-Index	0.678	0.620	0.782
	$\chi 2$ Value	1040.9	676.7	1141.7
	P-Value	0.000	0.000	0.000

Table 6.1 Moody's vs. Standard & Poor's (S&P): Comparisons of At Issuance Ratings

			Average Ranking[1]			Std. Dev. of Ranking[1]			Wilcoxon Signed-Rank Test	
		# of Issuances	Moody's	S&P	Gap[2]	Moody's	S&P	Gap	Signed-Rank Score	p value
Industrial Sector:	Communications	155	10.19	10.75	-0.56	3.62	3.04	1.25	-1528	<.0001
	Consumer	57	12.42	12.35	0.07	2.23	2.06	0.98	30	0.525
	Energy	80	9.55	10.30	-0.75	1.57	1.66	0.77	-888	<.0001
	Financial	39	9.36	9.59	-0.23	5.69	5.44	2.08	-18	0.264
	Industrial & Materials	107	13.89	13.01	0.88	1.89	1.78	1.04	1128	<.0001
	Other	17	9.59	10.06	-0.47	2.79	2.73	0.51	-18	0.008
Size of issuance:	< US$ 20 Million	22	7.91	7.91	0.00	5.08	4.41	2.60	5	0.796
	US$ 20 − US$ 200 Million	77	11.82	11.73	0.09	4.20	3.78	0.93	62	0.383
	US$ 200 − US$ 400 Million	165	12.83	12.50	0.33	2.57	2.03	1.24	1087	<.0001
	US$ 400 Million Plus	191	9.76	10.43	-0.67	2.87	2.60	1.14	-2991	<.0001
TOTAL		455	11.13	11.28	-0.15	3.52	3.00	1.33	-2955	0.033

Table 6.2 Moody's vs. Fitch: Comparisons of At Issuance Ratings

			Average Ranking[1]			Std. Dev. of Ranking[1]			Wilcoxon Signed-Rank Test	
		# of Issuances	Moody's	Fitch	Gap[2]	Moody's	Fitch	Gap	Signed-Rank Score	p value
Industrial Sector:	Communications	121	8.93	9.50	-0.56	2.90	2.89	1.16	-794	<.0001
	Consumer	58	11.98	11.84	0.14	2.10	2.31	1.15	48	0.256
	Energy	68	9.29	10.29	-1.00	1.52	1.52	0.49	-1073	<.0001
	Financial	27	7.37	8.41	-1.04	3.14	3.30	2.70	-76	0.009
	Industrial & Materials	74	13.93	12.46	1.47	1.77	1.99	1.02	1154	<.0001
	Other	6	9.33	9.50	-0.17	2.07	0.84	1.33	-2	1.000
Size of issuance:	< US$ 20 Million	17	9.41	10.35	-0.94	2.96	3.39	3.27	-16	0.223
	US$ 20 − US$ 200 Million	50	11.18	11.04	0.14	3.91	3.48	1.01	44	0.352
	US$ 200 − US$ 400 Million	106	12.58	12.05	0.54	2.34	2.20	1.46	743	<.0001
	US$ 400 Million Plus	181	9.07	9.60	-0.53	2.63	2.32	1.25	-2676	<.0001
TOTAL		354	10.44	10.57	-0.14	3.19	2.75	1.52	-1830	0.118

[1] While averages and standard deviations are not valid for ordinal scales they are presented here for informative purposes only.
[2]A positive gap indicates that Moody's has been more severe in its ratings; a negative gap indicates that Fitch has been more severe.

Table 6.3 Standard and Poor's (S&P) vs. Fitch: Comparisons of At Issuance Ratings

			Average Ranking[1]			Std. Dev. of Ranking[1]			Wilcoxon Signed-Rank Test	
		# of Issuances	S&P	Fitch	Gap[2]	S&P	Fitch	Gap	Signed-Rank Score	p value
Industrial Sector:	Communications	117	9.79	9.53	0.26	2.47	2.91	1.17	438	0.001
	Consumer	61	12.02	11.74	0.28	2.13	2.33	0.86	57	0.022
	Energy	68	10.24	10.38	-0.15	1.56	1.44	0.50	-28	0.027
	Financial	34	11.53	10.82	0.71	3.99	3.52	1.09	88	<.0001
	Industrial & Materials	87	13.23	12.60	0.63	1.77	1.69	0.82	546	<.0001
	Other	8	10.50	9.88	0.63	1.60	0.99	0.74	5	0.125
Size of issuance:	< US$ 20 Million	15	10.07	9.93	0.13	2.05	2.46	0.83	4	0.766
	US$ 20 − US$ 200 Million	75	12.23	11.77	0.45	3.17	2.95	0.70	254	<.0001
	US$ 200 − US$400 Million	114	12.20	11.87	0.33	1.83	2.10	1.12	354	0.001
	US$ 400 Million Plus	171	10.19	9.91	0.27	2.53	2.51	0.97	706	<.0001
TOTAL		375	11.20	10.88	0.32	2.67	2.66	0.97	3965	<.0001

1 While averages and standard deviations are not valid for ordinal scales they are presented here for informative purposes only.
[2]A positive gap indicates that S&P has been more severe in its ratings; a negative gap indicates that Fitch has been more severe.

Table 7 Pairwise Comparisons of Issues Rated by All Three Rating Agencies

			Average Ranking[1]			Std. Dev. of Ranking[1]			Wilcoxon Signed-Rank Test	
		# of Issuances	Rating Agency 1 (RA1)	Rating Agency 2 (RA2)	Gap[2]	Rating Agency 1 (RA1)	Rating Agency 2 (RA2)	Gap	Signed-Rank Score	p value
At Issuance	Moody's (RA1) vs. S&P (RA2)	308	10.58	10.88	-0.30	3.12	2.60	1.24	-3604	<.0001
	Moody's (RA1) vs. Fitch (RA2)	308	10.58	10.61	-0.03	3.12	2.68	1.45	-395	0.6798
	S&P (RA1) vs. Fitch (RA2)	308	10.88	10.61	0.27	2.60	2.68	1.00	2200	<.0001

[1] While averages and standard deviations are not valid for ordinal scales they are presented here for informative purposes only.
[2]A positive gap indicates that Rating Agency 1 has been more severe in its ratings; a negative gap indicates that Rating Agency 2 has been more severe.

Table 8 Pairwise Comparisons of Ratings at Issuance: 1994-2002 and 2003-2012

			Average Ranking[1]			Std. Dev. of Ranking[1]			Wilcoxon Signed-Rank Test	
		# of Issuances	Rating Agency 1 (RA1)	Rating Agency 2 (RA2)	Gap[2]	Rating Agency 1 (RA1)	Rating Agency 2 (RA2)	Gap	Signed-Rank Score	p value
1994−2002	Moody's (RA1) vs. S&P (RA2)	197	12.64	12.43	0.21	2.53	2.23	1.35	1085	0.0061
	Moody's (RA1) vs. Fitch (RA2)	117	12.39	12.11	0.28	1.95	2.10	1.70	554	0.0402
	S&P (RA1) vs. Fitch (RA2)	121	12.26	11.91	0.36	1.41	1.98	1.31	392	0.0039
2003−2012	Moody's (RA1)vs. S&P (RA2)	258	9.98	10.40	-0.42	3.73	3.22	1.25	-3400	<.0001
	Moody's (RA1) vs. Fitch (RA2)	237	9.47	9.81	-0.34	3.24	2.72	1.38	-2311	<.0001
	S&P (RA1) vs. Fitch (RA2)	254	10.70	10.39	0.31	2.96	2.80	0.75	1901	<.0001

[1] While averages and standard deviations are not valid for ordinal scales they are presented here for informative purposes only.
[2]A positive gap indicates that Rating Agency 1 has been more severe in its ratings; a negative gap indicates that Rating Agency 2 has been more severe.

Fig. 1 Number of New Issuances Rated per year by Rating Agency: 1994−2012

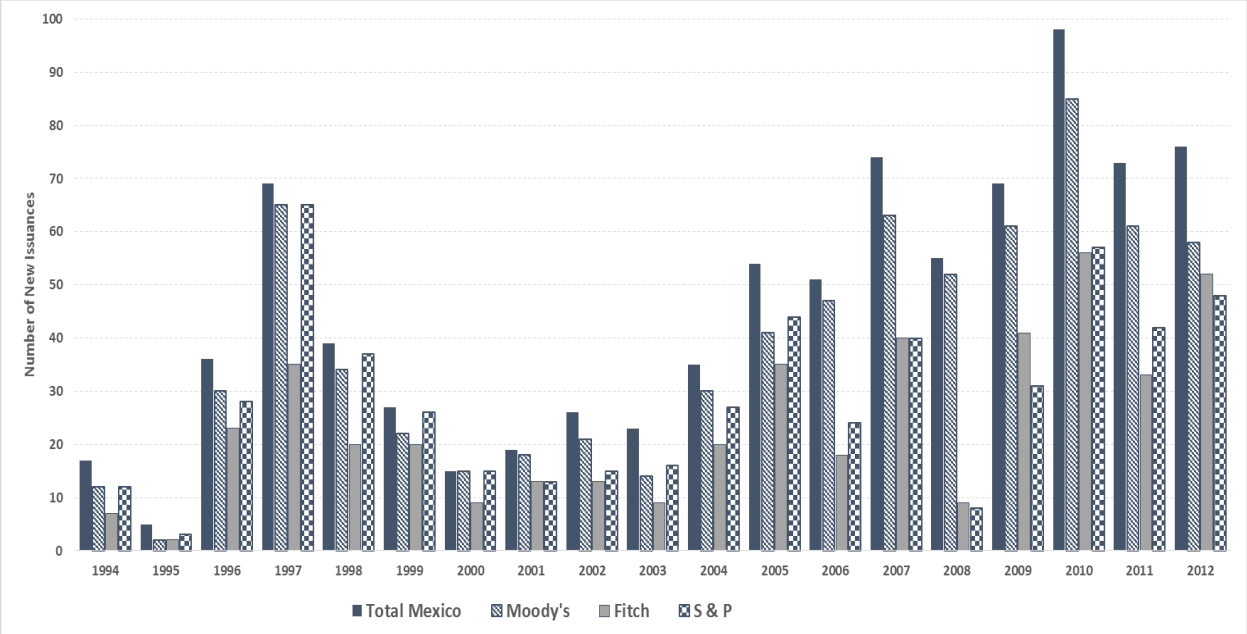

Fig. 2 Percent of New Issuances per Industrial Sector by Rating Agency: 1994−2012