FAUSTO HERNÁNDEZ-TRILLO
RICARDO SMITH-RAMÍREZ

Credit Ratings in the Presence of Bailout: The Case of Mexican Subnational Government Debt

Bond ratings have existed for nearly a century, and they have become a matter of public policy concern (Cavallo, Powell and Rigobón 2008). Debt issued by firms, sovereign countries, and subnational governments (SNGs) are regularly rated in industrial countries (Cantor and Packer 1995, 1996).[1] The rating history for less developed countries (LDCs) is shorter. International raters turned their attention to LDCs in the 1980s when agencies started rating LDC sovereign bonds in reaction to several international debt crises. As a result, literature on grading SNGs and sovereign bonds in industrial countries abounds, while for LDCs it is scarce.

Rating agencies have come under scrutiny in regard to their grading of industrial countries and LDCs. For example, the *Wall Street Journal* (2004) reported that credit ratings in China could be merely guesswork. In the case of sovereign credit ratings, there is a growing body of literature that casts doubt on their role (see, for example, Reinhart 2001 and 2002), especially after the Asian and Argentinean crises of 1997–98 and 2001, respectively. Others have attempted to refine the measurement of risk (Remolona, Scatigna, and Wu 2008; Alfonso 2003). More recently, the credibility of rating agencies has

Hernández-Trillo and Smith-Ramírez are with Centro de Investigación y Docencia Económicas (CIDE), Mexico City.

For helpful comments, we would like to thank participants in the seminars held at the Universidad de las Américas, Puebla; Banco de México; Ohio State University; University of Chicago Harris School Public Policy; Pontificia Universidad Católica del Perú, Lima; CIDE; University of California, Los Angeles; and especially those at the *Economia*—Latin American and Caribbean Economic Association LACEA seminar held in Bogotá, Colombia. We also would like to thank our discussants and the editors of this volume, Tito Cordella, Eduardo Cavallo, and Roberto Rigobón.

1. See Carleton and Lerner (1969) for pioneering work on SNGs.

been challenged due to the role of their evaluations in losses associated with the U.S. mortgage crisis.

In this paper, we target the rating technology used by agencies to grade SNGs in LDCs. By using data from the SNG bond market of Mexico, a country with a tradition of bailouts, we analyze how political and financial factors are weighted in the construction of ratings. This case exemplifies the situation in most Latin American countries.

Latin American governments have a long tradition of bailing out SNGs; Bevilaqua and Garcia (2002) document this phenomenon in Brazil, and Sanguinetti and others (2002) do likewise for Argentina. A high bailout probability raises at least two issues: the adequacy of the bond rating process and its purpose. Rating principles should take into account the many differences between industrial and LDC countries (Laulajainen 1999). Typically, developing countries have serious institutional and legal shortcomings (see Inter-American Development Bank 1997). Most relevant, they are very centralized, law enforcement is deficient (La Porta and others 1998), and most of them have just started fiscal decentralization reform, which in many cases has responded more to political pressure than to efficiency-enhancing purposes (see Díaz 2006; Giugalle, Korobow, and Webb 2001). These characteristics are important when rating bonds in their local currencies. For instance, Mexican SNGs are not allowed to issue debt denominated in foreign currency. Such differences call for different rating technologies than those used when rating entities within industrial countries, where many of the aforementioned shortcomings are not present.

Surprisingly, one of the largest states in Mexico, the State of Mexico, has been *continuously* bailed out since 1995; though this SNG is virtually bankrupt, it still has been assigned an investment grade.[2] Likewise, Sanguinetti and others (2002) report that the provincial government of La Rioja, Argentina, was bailed out several times previous to the 2001 crisis, and it still continues to receive an investment grading.[3] Bond ratings are meant to indicate the likelihood of default (Bhatia 2002).[4] If SNGs are to be bailed out anytime they

2. Reported by Black (2003). The grade assigned by Fitch in 2003 was BBB.

3. Sanguinetti, among others, argues that the Argentinean crisis was in part due to the fiscal permissiveness of SNGs in that country. For this reason, raters were questioned in Argentina.

4. It has been shown that these agencies specialize in gathering and processing financial information and are certified by screening agents who, in turn, are able to diversify their risky payoffs. In this setting, raters solve, at least in part, the informational asymmetry in capital markets, involving insiders possessing more accurate information about the true economic values of their firms or governments than outsiders. In turn, rating agencies gain from sharing their information. See Millon and Thakor (1985).

face financial problems, then their risk is passed on to the federal government. Thus SNG rates eventually would become similar to those of sovereign debt.[5] Is this happening in LDCs? If so, then the purpose of rating SNG debt may be arguable.

Rating agency results are puzzling in LDCs since, as pointed out before, they often give high rates to financially bankrupt SNGs. What, then, are agencies actually rating? Are they rating financial soundness or just probability of bailout? Do rating agencies foster market discipline in the presence of implicit guarantees, or do they tend to exacerbate moral hazard problems?[6] In this article, we try to answer these questions. Bailout events are most frequently the result of political negotiations. Therefore, we focus our analysis on the relevance, if any, that political factors have in the rating technology of three agencies: Standard and Poor's (S&P), Fitch, and Moody's. More specifically, we want to know whether the number of voters and political party in power matter; given Mexico's long bailout tradition, we expect they do. Additionally, we analyze how financial factors influence rates. Finally, unlike previous studies, we study the determinants of choosing a specific grading agency.

Our econometrics extend and modify the seminal methodology of Moon and Stotsky (1993) in that we consider data from three rating agencies instead of two, and we use a novel formulation of the Monte Carlo expectation maximization algorithm (see Wei and Tanner 1990) to circumvent the estimation of multidimensional integrals instead of using probability simulators.

Our results indicate that rating agencies differ slightly in how they weight relevant variables to assess the risk. Most notably, we found a strong negative correlation between SNG population, our proxy for number of voters, and debt risk. We interpret this as a "too big to fail" situation, that is, rating agencies consider that large entities, because of their political power, are more likely to be bailed out when facing financial problems (Hernández, Díaz, and Gamboa 2002). A second strong determinant of ratings is whether the party governing the country is also governing the entity under evaluation. When the parties are the same, debt risk decreases significantly. This is evidence that raters take into account the bailout phenomenon based on both the population of the SNG and political affinity between SNG and federal governments.[7]

5. In many cases, grades ran counter to the "sovereign ceiling" rule, as we will show later. See Durbin and Ng (2005).

6. The "market discipline" approach to subnational finance requires that moral hazard derived from the possibility of a central government bailout be made insignificantly small (Londero 2005).

7. Population has been interpreted as a political variable in the U.S. system of federal transfers under the New Deal. For a discussion, see Wallis (1998, 2001) and Fleck (2001).

These results are, to the best of our knowledge, novel in the bond rating literature.

The paper is organized as follows. The first section provides a brief description of Mexican intergovernmental relations and reviews the SNG debt environment in Mexico. The second section presents a discussion about the opacity of Mexican SNGs. In the third part, we present the model, describe the variables, and examine some descriptive statistics. The fourth section discusses the empirical results, followed by final remarks in the conclusion.

A Brief Overview of Mexico's Intergovernmental Relations and SNG Debt Regulation

Mexico is a federal republic composed of three levels of government: the central or federal government; 32 local entities, which consist of 31 states and the Federal District; and 2,477 municipalities (hereafter referred to as SNGs). Like many countries in the Latin-American region, Mexico is characterized by strong regional and state disparities. While the Federal District and the states of Mexico and Nuevo Léon produce about 40 percent of total GDP, Chiapas, Guerrero, Hidalgo, and Oaxaca generate a subtotal of only 6.8 percent of total GDP. Clearly, the southern region of the country is by far the poorest.

Mexico follows a revenue sharing system where the federal government collects main taxes, namely corporate and personal income taxes, value-added tax, and most excise taxes. These constitute 95 percent of total public sector tax revenue. Through a formula, 20 percent of this revenue is redistributed among states and municipalities. These net block transfers are known as *participaciones.* The main deficiencies identified in the system have been the local governments' lack of tax independence and the formula itself. Recently, decentralization efforts have been undertaken. However, this decentralization has not included the revenue side and instead concentrates on expenditures. Moreover, the process has been anarchic and has responded to political pressures and not to efficiency purposes (Hernández 1998).

The way SNG debt is regulated perhaps provides one of the most important explanations for its behavior (Ter-Minnasian 1999). For this reason, we now explain the Mexican case in more detail. First, SNG borrowing is regulated by the national constitution, which specifies that states can only borrow in pesos and solely for productive investment. The details for guaranteeing state credits are contained in the National Fiscal Coordination Law, which stipulates that these entities can borrow from commercial and development banks and by writing bonds to finance investment projects, subject to the previous authorization of the state congress.

Before the tequila crisis of 1994–95, when a unique political party dominated the country, SNG debt was virtually decided by the federal government in a unilateral manner by direct control of state governments (Díaz 2003). Later, as a consequence of the rapid democratization of the country, this control ended. The new situation allowed states to take advantage of the federal government's concerns about both the banking system—nearly bankrupt as a result of the tequila crisis—and states' abilities to deliver public services (Hernández 1998).

Bailouts were common before the tequila crisis, though the largest in Mexican history was extended in 1995.[8] As a consequence, virtually no commercial bank developed an institutional capacity to assess subnational lending. When the tequila crisis erupted, most states had high debt ratios and federal bailout occurred.

To correct the situation, the Mexican federal government faced the challenge of guaranteeing that bailouts would not occur in the future. This would allegedly be solved by imposing an ex ante market-based mechanism. So a new regulatory framework for debt management by local governments was introduced in 1997.[9]

States and creditors were induced to make their own trust arrangements in the collateralizing of debt with the block transfers and assuming all the legal risks involved, thus providing recourse for the federal government. A link was established between the risk of bank loans to SNGs and government credit rating.

Currently, credit ratings performed by reputable international agencies are published on a global scale. Bank regulators use these ratings to assign capital risk weight for loans provided to states and to municipalities. To control

8. For a review of bailout events in Mexico, see Hernandez, Díaz, and Gamboa (2002).

9. Firms or governments benefit from obtaining a good rating by lowering the cost of servicing the debt. Many studies of industrial countries have demonstrated empirically that this is generally the case, as they have gained greater acceptance in the market. Ratings also have been used in financial regulation because they simplify the task of prudential regulation (Cantor and Packer 1995). Thus, as in the Mexican case, regulators have adopted ratings-dependent rules.

agency shopping, two ratings are mandatory for regulation. In case of a large discrepancy, the capital risk weight of the lower rate applies. The National Securities Commission recognizes three rating agencies: Moody's, Standard and Poor's, and Fitch.

The main purpose of the regulation is to discipline SNG debt markets, especially in a new framework characterized by the absence of federal intervention. Financially weaker states and municipalities are likely to be priced or rationed out of the market, while stronger ones would see interest rates on their loans fall (Giugalle, Korobow, and Webb 2001).

Another important element in the new regulation is the registration of SNG loans with the federal government. Registration is conditioned on the borrowing state or municipality being current on the publication of its debt, the related fiscal statistics from the preceding year's final accounts, and on all of its debt service obligations toward the government's development banks. At the same time, in order to make that registration appealing, unregistered loans are automatically risk weighed by the regulators at 150 percent.

Several elements need to be considered to ensure the success of this type of regulation, including the market credibility of the federal government's commitment to not bail out defaulting SNGs, the quality of the enforcement of capital rules, and the quality and reliability of SNG fiscal information, as well as homogeneity in accounting standards.

As we pointed out previously, the largest state in Mexico has been *continuously* bailed out in the past. Furthermore, states and municipalities currently differ in their accounting standards, and not all of them publish their financial statements (ARegional 2004). These elements cast some doubt on the success of the new regulation. No new SNG default and therefore no bailout has occurred so far; however, unless more stringent oversight is exerted over SNGs, there is no guarantee that they will not occur in the future.

Are Mexican SNGs Opaque?

SNG fiscal information is like a black box in Mexico, mainly due to lack of an adequate institutional and legal framework and lack of accounting standards.[10] In general, rule of law in Mexico is poor (La Porta and others 1998). This

10. For example, for some municipalities the service of paving roads is registered in current expenditures, whereas for others it is treated as an investment (Hernández 1998).

TABLE 1. Relative Opacity
Kappa Index

Entity	*United States*	*Mexico*
Banks	0.30	0.27
Other sectors	0.45	0.36
States and municipalities		0.13

Source: Morgan (2002) for U.S. values; authors' calculations for Mexican values.

problem is greater at state and municipal levels, where transparency is nonexistent since governments are not required to publish their financial statements (Ugalde 2003).

Transparency issues should be taken into account when rating SNG bonds. Were SNGs transparent, there would be no need for a lender of last resort since fully transparent states could borrow at market rates that fairly reflected their risk. However, SNG transparency—and thus financial soundness—is more a matter of faith than fact in Mexico.

To discuss this point, we use Morgan's definition of relative opacity, which is framed in terms of disagreement between the major bond rating agencies—Fitch, S&P, and Moody's—when grading an entity and is used as a proxy for uncertainty (see Morgan 2002). The argument is this: if SNG risk is harder to observe, raters in the business of judging risk should disagree more over SNG bond issues than over other entities. As table 1 demonstrates, this is the case with Mexican SNGs. This table presents kappa statistics, which are used as a measure of disagreement in biometrics (Cohen 1968).[11] Kappa essentially locates raters along a spectrum between complete disagreement (kappa = 0) and complete agreement (kappa = 1).

Kappa is 0.13 for the whole set of Mexican SNGs—states and municipalities—rated by the three agencies, which suggests a strong disagreement. This figure worsens to 0.05 if only state governments are included. Some SNGs have applied only for two ratings. In this case, when the agencies are Fitch and S&P, the kappa is 0.24; when they are Fitch and Moody's, the figure is 0.17; and finally, when they are Moody's and S&P, the kappa indicator is 0.04. These figures suggest that SNGs are opaque according to Morgan's definition. U.S. SNGs rated by Moody's and Fitch have a kappa of 0.61, which suggests that these entities are not as opaque as those in Mexico.

11. Kappa = $(p_o - p_e)/(1 - p_e)$, where p_o is the observed percentage of graded bonds equally, and p_e is the expected percentage, given the current distribution of grades.

Ederington, Yawitz, and Roberts (1987) suggest three reasons for differing bond ratings. First, agencies may agree on the creditworthiness of a bond but apply different standards for a particular rating. Second, they may differ systematically in the factors they consider or the weights they attach to each factor. And third, due to the inherent subjectivity of the process, they may give different ratings for random reasons.

In this article, we expect to shed some light on which of these three explanations for disparities predominates when subnational entities in Mexico are rated.

Empirical Model and Estimation

A selectivity problem arises during the analysis of the determinants of SNG bond rating. This follows from the fact that ratings are observed only for those municipalities that have chosen to be rated rather than for all entities in the sample with outstanding debt.

As in Moon and Stotsky (1993), we treat this self-selection problem by developing a model in which we jointly analyze the determinants of the bond rating and the determinants of the decision to obtain a rating. Due to the short history of SNG bond rating in Mexico and in order to gather enough information for our study, we collected ratings from three agencies (Moody's, S&P, and Fitch) instead of the two (Moody's and S&P) used by Moon and Stotsky. Although estimating a three-agency model is more challenging, it has the advantage of expanding the scope of our conclusions, as we are now able to compare the rating technologies of more agencies. Additionally, by controlling for trivariate self-selection, we can consider in the analysis not only SNGs with three ratings but also those with only one or two ratings, as well as SNGs with no ratings but with outstanding debt.

We also examine jointly the determinants of the bond ratings for the three rating agencies. A joint estimation enables more efficient estimates by allowing free correlation between selection and rating equations. Allowing free correlation is important since rating decisions are not necessarily independent. Recall that an SNG needs at least two ratings in order to issue a bond registered with the Mexican treasury department, and we are considering three measures of credit risk obtained by the three agencies. Thus SNG administrators may show preferences for certain agencies if they believe these agencies have a less stringent rating procedure. Additionally, an entity may have incentives to obtain more than one or two ratings if by doing so it lowers the cost of its debt. The literature shows evidence that not only value of ratings but also the number of them influence the cost of debt (Ederington Yawitz, and Roberts 1987). Hence a multivariate framework applies.

The Model

Following the discussion above, the equation system to solve is:

(1)
$$
\begin{aligned}
y_s^* &= \mathbf{X}_s\boldsymbol{\beta}_s + \boldsymbol{\varepsilon}_s && \text{propensity to obtain S\&P's ratings}\\
w_s^* &= \mathbf{Z}_s\boldsymbol{\gamma}_s + \boldsymbol{\eta}_s && \text{S\&P's perceived riskiness}\\
y_f^* &= \mathbf{X}_f\boldsymbol{\beta}_f + \boldsymbol{\varepsilon}_f && \text{propensity to obtain Fitch's ratings}\\
w_f^* &= \mathbf{Z}_f\boldsymbol{\gamma}_f + \boldsymbol{\eta}_f && \text{Fitch's perceived riskiness}\\
y_m^* &= \mathbf{X}_m\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m && \text{propensity to obtain Moody's rating}\\
w_m^* &= \mathbf{Z}_m\boldsymbol{\gamma}_m + \boldsymbol{\eta}_m && \text{Moody's perceived riskiness,}
\end{aligned}
$$

where index $k = s, f, m$ refers to S&P, Fitch, and Moody's, respectively; matrices $\mathbf{X}_k$ and $\mathbf{Z}_k$ are matrices of explicatory variables; and $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$ are vectors of parameters to be estimated. The disturbance vector is assumed to be i.i.d. over entities according to the following six-dimensional normal distribution:

(2)
$$
\begin{pmatrix}\boldsymbol{\varepsilon}_{s,i}\\ \boldsymbol{\eta}_{s,i}\\ \boldsymbol{\varepsilon}_{f,i}\\ \boldsymbol{\eta}_{f,i}\\ \boldsymbol{\varepsilon}_{m,i}\\ \boldsymbol{\eta}_{m,i}\end{pmatrix} \sim \mathbf{N}_6\left(\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}
1 & \rho_{\varepsilon_s\eta_s} & \rho_{\varepsilon_s\varepsilon_f} & \rho_{\varepsilon_s\eta_f} & \rho_{\varepsilon_s\varepsilon_m} & \rho_{\varepsilon_s\eta_m}\\
\rho_{\varepsilon_s\eta_s} & 1 & \rho_{\eta_s\varepsilon_f} & \rho_{\eta_s\eta_f} & \rho_{\eta_s\varepsilon_m} & \rho_{\eta_s\eta_m}\\
\rho_{\varepsilon_s\varepsilon_f} & \rho_{\eta_s\varepsilon_f} & 1 & \rho_{\varepsilon_f\eta_f} & \rho_{\varepsilon_f\varepsilon_m} & \rho_{\varepsilon_f\eta_m}\\
\rho_{\varepsilon_s\eta_f} & \rho_{\eta_s\eta_f} & \rho_{\varepsilon_f\eta_f} & 1 & \rho_{\eta_f\varepsilon_m} & \rho_{\eta_f\eta_m}\\
\rho_{\varepsilon_s\varepsilon_m} & \rho_{\eta_s\varepsilon_m} & \rho_{\varepsilon_f\varepsilon_m} & \rho_{\eta_f\varepsilon_m} & 1 & \rho_{\varepsilon_m\eta_m}\\
\rho_{\varepsilon_s\eta_m} & \rho_{\eta_s\eta_m} & \rho_{\varepsilon_f\eta_m} & \rho_{\eta_f\eta_m} & \rho_{\varepsilon_m\eta_m} & 1
\end{bmatrix}\right),
$$

where $i = 1, \ldots, N$ and N is the sample size. Note that all observations contribute to the estimation of the correlation terms $\rho_{\varepsilon_j\varepsilon_k}$ $j, k = s, f, m.$ However, due to self-selection, only those SNGs that have received ratings from the respective agencies contribute to the estimation of terms $\rho_{\varepsilon_j\eta_k}$. Variable $y_{k,i}^*$ is not observable, but a binary counterpart, $y_{k,i}$, takes the value of 1 if $y^*_{k,i} > 0$ and takes the value of 0 otherwise. The observable counterpart of $w^*_{k,i}$ is categorical ordered so that

$$(3) \qquad w_{k,i} \begin{cases} l_{k,1} & \alpha_{k,1} < w^*_{k,i} \leq \alpha_{k,2} \\ l_{k,2} \quad \text{if} & \alpha_{k,2} < w^*_{k,i} \leq \alpha_{k,3} \\ \vdots & \vdots \\ l_{k,r} & \alpha_{k,r} < w^*_{k,i} \leq \alpha_{k,r+1} \end{cases},$$

where $l_{k,1} < l_{k,2} \cdots < l_{k,r}$ are consecutive integer values, $\alpha_{k,1} = -\infty$, $\alpha_{k,r+1} = \infty$, and thresholds $\alpha_{k,2} < \alpha_{k,3} < \ldots < \alpha_{k,r}$ are extra parameters to estimate. In our analysis, we have six categories for all agencies, that is, $r = 6$ with $l_{k,1} = 0$ and $l_{k,6} = 5 \ \forall \ k$ (see table 2). If $y_{k,i} = 0$, then $w_{k,i}$ does not exist, in accordance with the self-selection mechanism discussed above. Given the binary and categorical ordered nature of the observed counterparts of the dependent variables, parameter identification requires normalization of the diagonal elements in the disturbance covariance matrix as it is presented in equation 2. Additionally, identification of the coefficients γ_k in the perceived riskiness equations requires either setting to zero one of the thresholds in equation 3 for each equation or setting the intercept parameter in these equations equal to zero. We chose the first alternative and set $\alpha_{k,2} = 0$, $k = s, f, m$.

Model Specification, Data, and Description of Variables

In theory an entity decides to obtain a credit rating because it expects to save enough interest costs to outweigh the agency fee. Thus the level of debt may

TABLE 2. Equivalence between Ordinal and Qualitative Rates

	Qualitative rating by institution		
Ordinal rate	*S&P*	*Fitch*	*Moody's*
0	AA+, AA	AA	Aa2
1	AA–	AA–	Aa3
2	A+	A+	A1
3	A	A	A2
4	A–	A–, A3	A3
5	BB+, BB–	BBB+, BBB	Baa1, Bba1

Source: Authors' determinations.

be a good determinant of the propensity to be rated since the higher the debt, the greater the savings in interest costs (see Moon and Stotsky 1993).

Likewise, as in most previous literature, we include the total revenue of the entity, as it may represent a good proxy for the local income tax base, and it allows controlling for the size of the entity in terms of economic importance.

If large municipalities, in terms of population, perceive that they will be bailed out, they will then have strong incentives to be rated and obtain debt. We use population as a proxy for size importance in political terms, after controlling for economic size, since Hernández, Díaz, and Gamboa (2002) have shown that in the past more populated entities have been bailed out more favorably than less populated ones. This variable also has been discussed in the U.S. case. Wallis (1998, 2001) and Fleck (2001) maintain a debate about the political motive of using population during New Deal transfers to states. Since we use a log specification, the inclusion of total revenue and population rules out the possibility of adding revenue per capita as a regressor to avoid perfect collinearity. Nonetheless, during the estimation process, we tried using total revenue and revenue per capita alternatively; we detected neither significant qualitative nor quantitative differences in the final results except, of course, in the coefficients of the two regressors.

Finally, we control for political party, hypothesizing that the left-wing party has either less financial culture or dismisses market-based approaches with respect to obtaining debt. Thus dummies for the main political parties were included in the propensity equation.

Regarding the risk assessment equations, the major categories include political factors, some indicators of financial soundness including contingent liabilities, indicators of debt level, and economic indicators such as gross state product and its composition. Next we describe the variables considered in our analysis.

Again, population size in political terms is a variable that may affect rating behavior, in two ways in particular. First, as previously mentioned, political decisionmaking varies with the size of population. Hernández, Díaz, and Gamboa (2002) have shown that this variable is a good proxy for the "too big to fail" hypothesis for state bailouts. In this sense, the larger the entity, the higher the number of political votes it has. Second, population is important as a measure of tax base in Mexico. This may be different in advanced economies where smaller municipalities tend to be mostly residential, while larger municipalities tend to have a more substantial industrial base and a more diverse population. In contrast, in LDCs—and Mexico is no exception—

small municipalities tend to be more rural and thus less subject to being taxed. Complementarily, to control for economic size, we include the entity's total revenues. This is necessary since there may be municipalities that are small in terms of population but large in terms of economic importance.[12]

We include a dummy with the value one when the political party in control of the municipal government is the same as that controlling the federal government, and with the value zero otherwise. As already mentioned, we expect that raters assign a greater probability of bailout to entities that share political affinity with the central government, which therefore translates to a better risk grade.

For financial soundness, we choose several variables previously used in the literature (see, for example, Ederington, Yawitz, and Roberts 1987; Cantor and Packer 1996). We include the ratio of an entity's own revenues to total revenue for two reasons. First, it reflects the flexibility an entity has to absorb a shock; and second, federal transfers to total revenue reflect how compromised the transfer is beforehand. With respect to debt, we use debt-to-income ratio. Mexican law requires that all new debt must be used in public investment. Thus one would expect that higher levels of fiscal responsibility imply larger amounts of investment; for this reason, we also include the investment-to-total-expenditure ratio.

Regarding the functional form assumed for the model, we follow Moon and Stotsky (1993) and use the log form of all the continuous regressors. The data set contains information from 149 urban municipalities for the year 2001, 148 municipalities for the year 2002, and 147 municipalities for the year 2003.[13] Descriptive statistics for the data are presented in table 3. We obtain the financial and political variables from the Municipal Information System of the National Institute of Statistics (2003).

Estimation Approach

In contrast to the selection problems involving continuous or limited dependent response variables, the responses to the debt risk according to the different agencies are not observed at all in our problem. All that we know about these responses is a discrete ordinal manifestation in the ratings. Thus a selectivity-

12. The typical example of this in Mexico is San Pedro in the state of Nuevo León.

13. Remember that the regulation is biased toward the largest 150 municipalities in Mexico, and that the remaining municipalities were virtually excluded from credit markets, as argued above (see Hernández 1998).

TABLE 3. Descriptive Statistics of Dependent and Explanatory Variables

Binary dependent variables		*Sum*	
S&P	Entity rated by S&P in the period 2001–03 (yes = 1)	96	
Fitch	Entity rated by Fitch in the period 2001–03 (yes = 1)	74	
Moody's	Entity rated by Moody's in the period 2001–03 (yes = 1)	40	
Dummy explanatory variables		*Sum*	
PRD	Entity administered by the Partido Revolucionario Democrático (yes = 1)	45	
PRI	Entity administered by the Partido Revolucionario Institutional (yes = 1)	148	
COA	Entity administered by a coalition party (yes = 1)	60	
PAN	Entity administered by the Partido Acción Nacional (yes = 1)	191	
A	Entity administered by the same party as federal government (yes = 1)	191	
Continuous explanatory variables [a]		*Mean*	*Std. deviation*
POP	2000 population ($\times 10^5$)	3.3	3.1
TI	Total annual income (U.S.$ $\times 10^8$)	25.7	31.7
O_T	Own-to-total revenue ratio	0.23	0.12
D_I	Debt-to-income ratio	0.12	0.17
Debt	Total debt (U.S.$ $\times 10^6$)	20.8	44.8
P_D	Per capita debt (U.S.$ $\times 10^3$)	0.58	0.90
I_G	Investment-to-total-expenditure ratio	0.23	0.13

Source: National Institute of Statistics (2003).
a. The $\log_{10}$ form of the continuous explanatory variables was used in the estimation.

corrected Heckman-type estimator cannot be calculated since least squares cannot be applied on an unobserved variable in the second stage of Heckman's procedure. Therefore we use a full information maximum likelihood (FIML) approach.

It is well known that the main problem when using FIML to estimate equation systems involving latent variables is the presence of high dimensional integrals in the likelihood function, the highest possible order of integration being equal to the number of latent variables in the system. In contrast to Moon and Stotsky (1993), who use the probability simulator of Börsch-Supan and Hajivassiliou (1993), we deal with this issue by formulating a Monte Carlo expectation maximization (MCEM) algorithm. The main advantages of the MCEM approach are its robustness both to the selection of starting values and to fragile identification (Natarajan and others 2000).

To get a feel for how the MCEM method works, consider the following many-to-one mapping, $\mathbf{z} \in Z \rightarrow \mathbf{y} = y(z) \in Y$. In other words, $\mathbf{z}$ is only known to lie in $Z(y)$, and the subset of Z is determined by the equation $\mathbf{y} = y(\mathbf{z})$, where $\mathbf{y}$ is the observed data variables y_k and w_k in our case, and $\mathbf{z}$ is the unobserved information, our y_k^* and w_k^* variables. Thus the complete data is $\mathbf{x} = (\mathbf{y}, \mathbf{z})$, and the log-likelihood of the observed information is

$$\ell\left(\theta\middle|\mathbf{y}\right) = \ln L\left(\theta\middle|\mathbf{y}\right) = \ln \int_{Z(y)} L\left(\theta\middle|\mathbf{x}\right) d\mathbf{z}. \tag{4}$$

Hence the multidimensional integration problem appears when we try to exclude the unobserved information by integration. Instead of trying to solve equation 4 directly, the expectation maximization (EM) algorithm focuses on the complete information log-likelihood $\ell^c(\theta|\mathbf{x})$, and maximizes $E[\ell^c(\theta|\mathbf{x})]$ by executing two steps iteratively (Dempster, Laird, and Rubin 1977). The first one is the so-called expectation step (E-step), which computes $Q(\theta|\theta^{(m)}, \mathbf{y}) = E[\ell^c(\theta|\mathbf{x})]$ at iteration m+1. The term $E[\ell^c\ (\theta|\mathbf{x})]$ is the expectation of the complete information log-likelihood conditional on the observed information, provided that the conditional density $f\ (\mathbf{x}|\mathbf{y}, \theta^{(m)})$ is known. The E-step is followed by the maximization step (M-step), which maximizes $Q(\theta|\theta^{(m)}, \mathbf{y})$ to find $\theta^{(m+1)}$. Then the procedure is repeated until convergence is attained.

The Monte Carlo version of the EM algorithm avoids troublesome computations in the E-step by imputing the unobserved information by Gibbs sampling (Casella and George 1992), conditional on what is observed and on distribution assumptions. In this approach, the term $Q(\theta|\theta^{(m)}, \mathbf{y})$ is approximated by the mean $\frac{1}{K}\sum_{k=1}^{K} Q(\theta, \mathbf{z}^{(k)}|\mathbf{y})$, where the $\mathbf{z}^{(k)}$ are random samples from $f(\mathbf{x}|\theta^{(m)}, \mathbf{y})$. The formulation of an MCEM algorithm for estimating equation system 1 is presented in appendix A, and the information matrix was obtained using Louis's identity (Louis 1982; see appendix B).

Determinants of Credit Ratings

Determinants of the Rating Propensity

Estimation results for the whole set of parameters in the model are given in tables 4 and 5. We dropped the dummy representing the left-wing political party, Partido Revolucionario Democrático (PRD), from the regression in order to compare the impact of political orientation on the propensity to be rated. Tables 6 and 7 show the marginal effects of the explanatory variables on propensity-to-be-rated and rating equations, respectively. As it is well known, direct discussion of parameter estimates can be misleading in nonlinear models

TABLE 4. Model Estimates[a]

		S&P		*Fitch*		*Moody's*	
Equation	*Variable*	*Estimate*	*Std. error*	*Estimate*	*Std. error*	*Estimate*	*Std. error*
Propensity to be rated	Constant	–1.4967***	0.3798	–1.0241***	0.3944	–4.9557***	0.4660
	PRI	0.3496	0.2497	0.4515**	0.2033	3.2767***	0.4196
	Coalition	0.5976**	0.2941	0.3982	0.2815	3.4165***	0.4330
	PAN	0.8437***	0.2611	0.9766***	0.2011	3.4630***	0.4176
	POP	1.8514***	0.3395	1.6517***	0.3202	1.0679**	0.4313
	TI	–0.0251	0.2197	–0.4905**	0.2361	0.0638	0.2566
	O_T	1.1772***	0.2541	0.7546***	0.2413	–0.0828	0.2801
	D_I	–0.0072	0.0573	0.1878***	0.0512	0.2601***	0.0633
Rating	Constant	0.9318	0.8961	0.8772	1.0615	1.9660	1.3718
	A	–0.7958***	0.2494	–0.7496**	0.3714	–1.9033***	0.3383
	POP	–2.3515***	0.6811	–1.7098*	0.9723	–1.5152	1.8990
	TI	0.2692	0.3831	0.1023	0.5397	0.4503	1.0746
	O_T	–2.5806***	0.8021	–2.8562***	0.8788	–2.3790***	0.9092
	D_I	0.1255	0.0925	0.0378	0.1409	0.1500	0.1873
	I_G	–1.0191**	0.4320	–1.2384**	0.5175	–1.5604**	0.6599

*Statistically significant at the 10 percent level; ** statistically significant at the 5 percent level; *** statistically significant at the 1 percent level.

a. Test for model significance (restricted model: slopes are all zero) chi-squared = 710.4315, gl = 39 ($p < 0.01$). For abbreviations, see table 3.

TABLE 5. Thresholds and Covariance Matrix

	S&P		*Fitch*		*Moody's*	
Thresholds	*Estimate*	*Std. error*	*Estimate*	*Std. error*	*Estimate*	*Std. error*
$\alpha_{k,3}$	0.5030***	0.1242	0.8303***	0.1732	2.6975***	0.3094
$\alpha_{k,4}$	1.4025***	0.1205	1.8727***	0.1200	3.0433***	0.2239
$\alpha_{k,5}$	2.0564***	0.1146	2.4665***	0.1385	3.8627***	0.3222
$\alpha_{k,6}$	2.9212***	0.2337	3.5228***	0.2850	4.5533***	0.2924
Covariance matrix	*Estimate*	*Std. error*	*Covariance matrix*		*Estimate*	*Std. error*
$\rho_{\epsilon_s \eta_s}$	–0.2675	0.1898	$\rho_{\eta_s \eta_m}$		0.7370***	0.1229
$\rho_{\epsilon_s \epsilon_f}$	0.6536***	0.0387	$\rho_{\epsilon_f \eta_f}$		–0.1022	0.5946
$\rho_{\epsilon_s \eta_f}$	–0.0745	0.3518	$\rho_{\epsilon_f \epsilon_m}$		0.0442	0.0778
$\rho_{\epsilon_s \epsilon_m}$	0.2840***	0.0635	$\rho_{\epsilon_f \eta_m}$		–0.1963	0.4274
$\rho_{\epsilon_s \eta_m}$	–0.4171	0.4097	$\rho_{\eta_f \epsilon_m}$		–0.0662	0.1986
$\rho_{\eta_s \epsilon_f}$	0.1288	0.2232	$\rho_{\eta_f \eta_m}$		0.6897***	0.2519
$\rho_{\eta_s \eta_f}$	0.6351***	0.1291	$\rho_{\epsilon_m \eta_m}$		0.0218	0.3303
$\rho_{\epsilon_m \eta_m}$	–0.1191	0.1832				

***Statistically significant at the 1 percent level.

TABLE 6. Marginal Effects for the Propensity-to-Be-Rated Equations[a]

	S&P		*Fitch*		*Moody's*	
Variable	*Estimate*	*Std. error*	*Estimate*	*Std. error*	*Estimate*	*Std. error*
PRI	0.0553	0.0354	0.0611**	0.0249	0.0815***	0.0164
Coalition	0.1052**	0.0482	0.0521	0.0387	0.1021***	0.0288
PAN	0.1633***	0.0400	0.1771***	0.0290	0.1097***	0.0168
POP	0.3873***	0.0703	0.3378***	0.0658	0.1498**	0.0633
TI	−0.0053	0.0460	−0.1003**	0.0485	0.0090	0.0360
O_T	0.2462	0.0542	0.1543***	0.0508	−0.0116	0.0393
D_I	−0.0015***	0.0120	0.0384***	0.0107	0.0365***	0.0094

Statistically significant at the 5 percent level; * statistically significant at the 1 percent level.

a. For abbreviations, see table 3.

since they measure the impact of the regressors on latent dependent variables, which might have an intuitive meaning but not a definite one (Greene 2000). Therefore we focus our discussion on marginal effects, which estimate the effect of regressors on the observed counterparts of the dependent variables for the sample under study. For the particular case of the propensity-to-be-rated equation, the marginal effect accounts for the change in the probability that an entity requests to be rated as a result of a change in the respective regressor. Marginal effects are calculated for each observation; we report sample averages and standard errors calculated by the delta method.

It turns out that political orientation is important. As observed, the propensity to request a rate increases as with the shift from the left- to the right-wing preferences. Thus it is the Partido Acción Nacional (PAN), the rightist party, that demonstrates the highest propensity. According to table 6, ceteris paribus, a municipality governed by the PAN shows a probability to be rated by S&P 16 percentage points (pp) higher than one ruled by the PRD, the leftist party. This figure climbs to approximately 18 pp for Fitch and decreases to 11 pp for Moody's.

This result indicates that entities governed by the PAN are the most willing to obtain a grade, a finding that makes sense since the PAN is associated with local entrepreneurs, a group with more financial culture than other constituencies (Cabrero 2004).

Another significant variable that explains propensity to be rated is municipality size, measured in population terms. According to table 6, if municipality A has twice the population of municipality B, then the probability that A asks for the services of S&P would be about 11 pp higher than it would be for B.[14]

14. To get this figure, multiply the corresponding marginal effect by $\log_{10}(2) \approx 0.3$.

The respective figures for Fitch and Moody's are 10 pp and 4 pp, all of them significant at the usual levels of significance.

It can be noted that, aside from political preferences, population is the most important variable in explaining the decision to be rated. This suggests the ex ante existence of a self-selection mechanism, where larger municipalities select themselves into the rating process. Regarding financial factors, ratio of own to total revenue is important among those that choose S&P and Fitch, while ratio of debt to total income is important among those that choose Fitch and Moody's.

Determinants of the Rating

Overall, the estimates support the arguments presented in this article, namely, that population, political affinity with the federal government, the ratio of own to total revenues, and the investment variable influence the grade positively. Coefficient signs are negative because we assign a lower risk to higher grades (see table 2). Conversely, the ratio of debt to total income affects the grade negatively.

It can be seen that political variables are important for rating agencies. On the one hand, as discussed earlier, population size is important probably because it is politically more costly not to rescue a large entity. On the other hand, the high significance of the dummy for political affinity is evidence that raters allocate a higher rate to those entities having a higher bailout probability, that is, entities administrated by the party holding federal office.[15]

For a discussion based on probabilities, table 7 presents the marginal effects of regressors on the probabilities of receiving a given grade 0 to 5 as described in table 2, conditional on the SNG requesting a rating.

The regressors that provide statistically significant marginal effects in the S&P rating equation are political affinity, population, ratio of own to total revenue, and ratio of debt to total revenue. Marginal effects for population indicate that a rise in population size shifts the probability distribution from lower to higher grades. In particular, a 10 percent average rise in population brings a 2 pp average increase in the probability of receiving an AA or AA^+ grade and 0.4 pp increase in the probability of receiving an AA^- from S&P, with simultaneous reductions of 1.0 pp and 0.8 pp in the probability of being rated with A^- or BB, respectively. Although an increase in population favors

15. Although, for the period covered in our analysis, no changes in the federal government occurred, local governments did change. This happens because federal elections may take place at different dates than municipal ones.

TABLE 7. Marginal Effects for the Rating Equations[a]

		S&P		Fitch		Moody's	
Variable	*Rate*[b]	*Estimate*	*Std. error*	*Estimate*	*Std. error*	*Estimate*	*Std. error*
TI	0	−0.0627	0.0904	−0.0072	0.0470	−0.0178	0.0456
	1	−0.0149	0.0216	−0.0079	0.0530	−0.0829	0.1928
	2	0.0000	0.0061	−0.0053	0.0374	0.0002	0.0044
	3	0.0209	0.0302	0.0026	0.0165	0.0102	0.0248
	4	0.0318	0.0461	0.0094	0.0627	0.0172	0.0393
	5	0.0250	0.0372	0.0084	0.0583	0.0732	0.1749
0_T	0	0.5662***	0.1826	0.2955***	0.0991	0.0943	0.0689
	1	0.1315***	0.0483	0.3337***	0.1175	0.4389***	0.1583
	2	−0.0055	0.0547	0.2398**	0.1102	−0.0011	0.0224
	3	−0.1885**	0.0750	−0.0999	0.0655	−0.0537	0.0542
	4	−0.2831***	0.1098	−0.3961***	0.1494	−0.0911	0.0756
	5	−0.2207**	0.1051	−0.3730**	0.1482	−0.3873***	0.1488
D_I	0	−0.0294	0.0216	−0.0053	0.0126	−0.0058	0.0093
	1	−0.0070	0.0052	−0.0061	0.0143	−0.0269	0.0343
	2	0	0.0028	−0.0046	0.0106	0.0001	0.0014
	3	0.0098	0.0074	0.0017	0.0044	0.0033	0.0063
	4	0.0149	0.0115	0.0073	0.0170	0.0056	0.0078
	5	0.0118	0.0097	0.0071	0.0162	0.0237	0.0299
I_G	0	0.2409**	0.0987	0.1305**	0.0620	0.0619	0.0386
	1	0.0577**	0.0281	0.1476**	0.0650	0.2880**	0.1439
	2	0.0005	0.0232	0.1063*	0.0610	−0.0007	0.0145
	3	−0.0801**	0.0390	−0.0440	0.0285	−0.0353	0.0311
	4	−0.1224*	0.0626	−0.1751**	0.0867	−0.0598	0.0565
	5	−0.0966*	0.0526	−0.1652*	0.0866	−0.2542*	0.1317

*Statistically significant at the 10 percent level; ** statistically significant at the 5 percent level; *** statistically significant at the 1 percent level.

a. For abbreviations, see table 3.

b. For equivalence between ordinal and qualitative rates, see table 2.

the probability of obtaining a better grade from Fitch as well, the changes in the distribution of that probability differ from S&P. Thus a 10 percent increase in population implies a 0.9 pp reduction in the probability of getting an A^- or BB rating and, similarly, 0.6 pp increase in the probability of obtaining an A^+, AA^-, or AA^+. In other words, changes in population tend to have a more uniform impact across the rates for Fitch, while for S&P they tend to affect the lowest and highest rates preferentially. On the other hand, population tested not significant in the Moody's rating equation.

Regarding political affinity, entities governed by the same party as the executive branch have a 14 pp higher probability of getting an AA^+ rating, and a 6 pp and 8 pp lower probability of obtaining an A^- and BB rating, respectively, from S&P than those governed by a different party. Corresponding results for Fitch are 6 pp, 9 pp, and 10 pp, and for Moody's, 3 pp, 14 pp, and 36 pp. Again, although the impacts of these determinants may seem to affect agencies in a similar qualitative way, their effects differ quantitatively as they relocate rate probabilities differently across agencies (see table 7).

Opacity

We have already demonstrated that agencies seem to take into account the same group of variables when constructing a grade. However, this condition is not sufficient to ensure that different agencies will grant the same grade to a single municipality. Agencies might consider the same factors, but they could weight them differently. In what follows, we test for SNG opacity by examining whether raters weight the factors in the same way when constructing a grade.

Direct examination of the sample indicates that among those entities rated by both S&P and Fitch, in only 60 percent of the cases did the two agencies grant the same grade to a particular entity. The proportion is smaller, 44 percent, among those rated by Fitch and Moody's, and only 38 percent among those rated by both S&P and Moody's.

In order to perform a statistical test to detect weighting differences across raters, we compare the marginal effects obtained for the rating equations. Three Wald tests comparing the estimates of the rating agencies by pairs showed high statistical differences between S&P and Moody's chi squared value—148.1, $p < 0.01$—and between Fitch and Moody's chi squared value—78.80, $p < 0.01$. Smaller but still significant differences were detected between S&P and Fitch's chi squared: 19.98, $p < 0.05$.

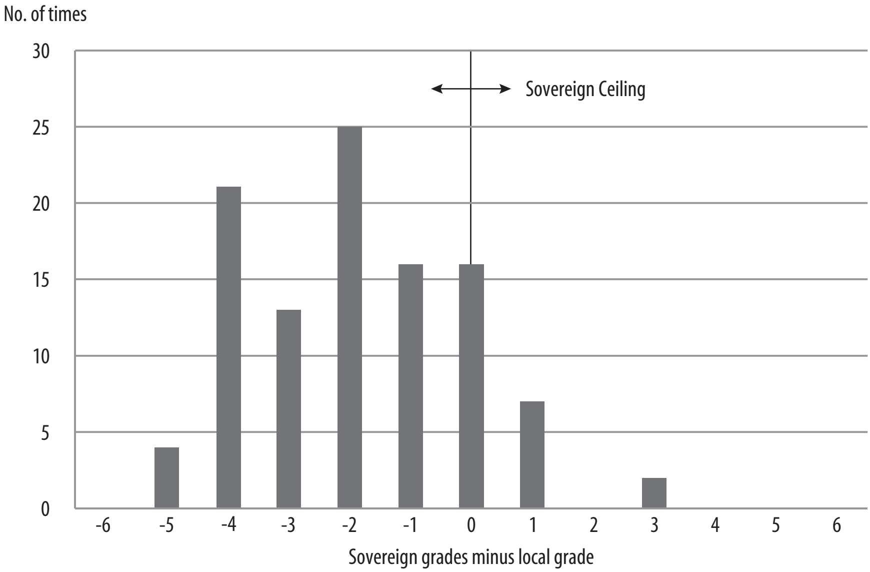
Overall, these results indicate that raters weigh factors differently during their rating process, which implies there is a high likelihood that they generate different rates even for the same municipality. This is consistent with the kappa analysis presented earlier.

Is There a Violation of "Sovereign Ceiling" Rule?

Figures 1 and 2 show histograms for the differences between SNG and sovereign ratings for S&P and Moody's, respectively. The differences were calculated as sovereign grade minus local grade.[16] As it may be noted, grades ran counter to the sovereign ceiling rule. Some have argued that this damages

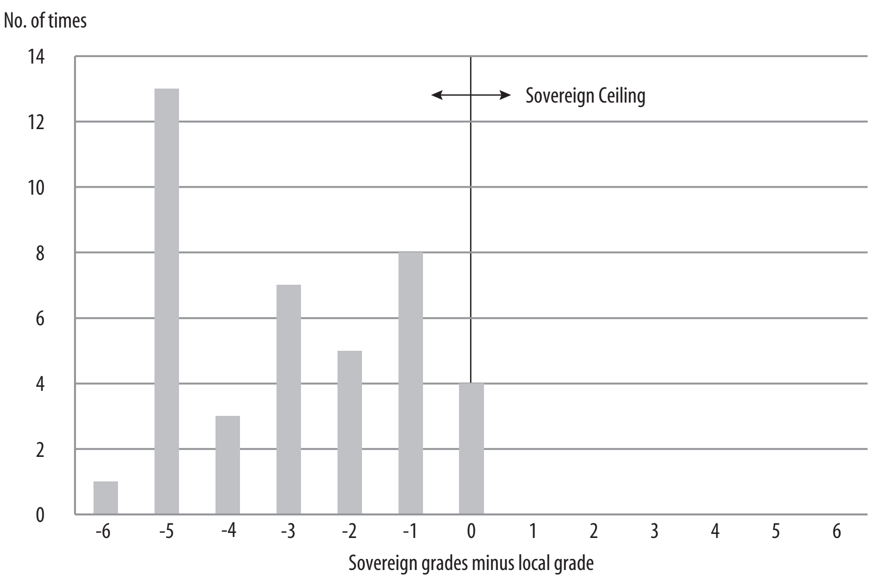
16. In contrast to table 2, where we use a 0–5 rating scale, for illustration purposes we use a 0–7 scale to construct these graphs.

FIGURE 1. S&P Distribution of Sovereign Ceilings, 2001–04[a]



Source: Authors' calculations.
a. In contrast to table 2, where we use a 0–5 rating scale, for illustration purposes we use a 0–7 scale to construct these graphs.

FIGURE 2. Moody's Distribution of Sovereign Ceilings, 2001–04[a]



Source: Authors' calculations.
a. In contrast to table 2, where we use a 0–5 rating scale, for illustration purposes we use a 0–7 scale to construct these graphs.

the credibility of raters since no domestic firm should get a better rate than its government (see Durbin and Ng 2005). Our econometric analysis showed that bailout probability is heavily weighted in the construction of SNGs' rates. On this basis, we would have expected the sovereign ceiling rule to hold for our sample since, in a country with a bailout tradition, the lender of last resort is the federal government. Therefore, the failure to conform to the sovereign ceiling rule, as illustrated in figures 1 and 2, is surprising. This contradiction indicates a disconnection between how SNG and sovereign rates are generated, something that should not happen in a country with a long history of bailout. In our opinion, this disconnect challenges the credibility of a market-based regulation.

In the introduction of this article, we mentioned that bond raters have been under scrutiny, especially after the crises in the nineties. Additionally, we noted that the operation of this market in less developed countries has not been studied, despite the fact that some doubts about its performance have been expressed (see the Chinese example at the beginning of this paper). Our results suggest that the puzzling grades often observed in LDCs could be explained if one considers not only financial factors in the construction of a rating but also political issues. We have proved that in a country with a history of bailout and opacity, such as Mexico, political variables become important in explaining the grade assigned to the debt of subnational governments, a fact that may undermine the credibility of a market-based regulatory framework.

Conclusions

In this paper, we studied both the determinants of the decision to be rated and the ratings for SNG debt in Mexico, a prominent LDC. One of the main findings is that not only financial but also political factors matter. We showed that population size is a strong determinant of debt rating. In a country with a long bailout history, this result supports our "too big to fail" hypothesis. First, large entities select themselves to be rated and so to obtain new debt because they know that they have political power; second, raters know that the probability that the federal government will bail out large entities is high. We also showed that political closeness between local and federal governments is important: rating agencies give a better grade to those entities being governed by the same party as the national executive branch. These outcomes challenge the purpose of rating subnational debt in LDCs with a bailout tradition, since the market may assess the risk of these entities as equivalent or superior to that of sovereign instruments.

Mexico has implemented new legislation for the SNG debt market, with the goal of increasing the transparency of the market and ruling out debt bailouts. According to our results, which show a high relevance of the bailout probability on ratings, it seems that bond rating agencies are not yet convinced of the success of such legislation. It is apparent that ratings methodologies take time to evolve, and, for the Mexican case at least, they continue echoing the market opacity and bailout tradition of the country. Mexican regulation in this sense needs to be revised to foster its credibility.

Appendix A. The MCEM Algorithm

Let **y** be a matrix containing all the observed information. The complete information log-likelihood function for the six-equation system discussed in the text is standard and can be written as the sum of the contributions from eight different regimes. The regimes are represented by the subsample receiving no grading; the potential three subsamples being graded by a single agency $k = s, f,$ or $m;$ the potential three subsamples being graded by two agencies; and the subsample receiving grades from the all three agencies. The corresponding contributions from the $j = 1, \ldots, 8$ regimes to the likelihood are

—regime $j = 1$: $y_{m,i} = y_{s,i} = y_{f,i} = 0$

$$\ell_j^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j \middle| \mathbf{y}\right) = -\frac{3n_j}{2}\ln\left(2\pi\right) - \frac{n_j}{2}\ln\left|\boldsymbol{\Omega}_j\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\sum_i \boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\right)$$

—regimes $j = 2$: $y_{m,i} = 1$; $y_{s,i} = y_{f,i} = 0$; $j = 3$: $y_{s,i} = 1$; $y_{m,i} = y_{f,i} = 0$; and $j = 4$: $y_{f,i} = 1$; $y_{m,i} = y_{s,i} = 0$

$$\ell_j^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j \middle| \mathbf{y}\right) = -2n_j \ln\left(2\pi\right) - \frac{n_j}{2}\ln\left|\boldsymbol{\Omega}_j\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\sum_i \boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\right) \quad j = 2, 3, 4$$

—regimes $j = 5$: $y_{m,i} = y_{s,i} = 1$; $y_{f,i} = 0$; $j = 6$: $y_{m,i} = y_{f,i} = 1$; $y_{s,i} = 0$; and $j = 7$: $y_{f,i} = y_{s,i} = 1$; $y_{m,i} = 0$

$$\ell_j^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j \middle| \mathbf{y}\right) = -\frac{5n_j}{2}\ln\left(2\pi\right) - \frac{n_j}{2}\ln\left|\boldsymbol{\Omega}_j\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\sum_i \boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\right) \quad j = 5, 6, 7$$

—regime $j = 8$: $y_{m,i} = y_{s,i} = y_{f,i} = 1$

$$\ell_j^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j \middle| \mathbf{y}\right) = -3n_j \ln\left(2\pi\right) - \frac{n_j}{2}\ln\left|\boldsymbol{\Omega}_j\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\sum_i \boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\right) \quad j = 8.$$

Thus

$$\ell^c\left(\boldsymbol{\theta}, \boldsymbol{\Omega}\middle|\mathbf{y}\right) = \sum_{j=1}^{8} \ell_j^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j\middle|\mathbf{y}\right), \tag{5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_m\ \boldsymbol{\gamma}_m\ \boldsymbol{\beta}_s\ \boldsymbol{\gamma}_s\ \boldsymbol{\beta}_f\ \boldsymbol{\gamma}_f)'$, $\boldsymbol{\theta}_j$ contains the components of $\boldsymbol{\theta}$ present in the equations solved for entities in regime j, $\boldsymbol{\Omega}_j$ is the covariance matrix of the disturbance terms associated to those equations j, n_j is the number of observations in regime j, and $\sum_j n_j = N$, the sample size.

E-Step

The expectation of expression 5 above, conditional on observed information and distribution assumptions, can be written as

$$E\left[\ell^c\left(\boldsymbol{\theta}_j, \boldsymbol{\Omega}_j\middle|\mathbf{y}\right)\right] = -\left[\frac{3}{2}n_1 + 2\left(n_2 + n_3 + n_4\right) + \frac{5}{2}\left(n_5 + n_6 + n_7\right) + 3n_8\right]\ln\left(2\pi\right) - \frac{1}{2}\sum_j n_j \ln\left|\boldsymbol{\Omega}_j\right| - \frac{1}{2}\sum_j \mathrm{tr}\left(\boldsymbol{\Omega}_j^{-1}\sum_i E\left[\boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\right]\right).$$

The E-step at iteration $m + 1$ requires the calculation of

$$\mathbf{Q}_{ji}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(m)}, \boldsymbol{\Omega}_j^{(m)}, \mathbf{y}\right) = E\left[\boldsymbol{\varepsilon}_{ji}\boldsymbol{\varepsilon}'_{ji}\middle|\boldsymbol{\theta}^{(m)}, \boldsymbol{\Omega}_j^{(m)}, \mathbf{y}\right]$$

$$= \sigma_{ji}^{2(m)} + \begin{pmatrix} \mu_{y^*_{m,i}}^{(m)} - \mathbf{X}_{m,i}\boldsymbol{\beta}_m \\ \mu_{w^*_{m,i}}^{(m)} - \mathbf{Z}_{m,i}\boldsymbol{\gamma}_m \\ \mu_{y^*_{s,i}}^{(m)} - \mathbf{X}_{s,i}\boldsymbol{\beta}_s \\ \mu_{w^*_{s,i}}^{(m)} - \mathbf{Z}_{s,i}\boldsymbol{\gamma}_s \\ \mu_{y^*_{f,i}}^{(m)} - \mathbf{X}_{f,i}\boldsymbol{\beta}_f \\ \mu_{w^*_{f,i}}^{(m)} - \mathbf{Z}_{f,i}\boldsymbol{\gamma}_f \end{pmatrix}\begin{pmatrix} \mu_{y^*_{m,i}}^{(m)} - \mathbf{X}_{m,i}\boldsymbol{\beta}_m \\ \mu_{w^*_{m,i}}^{(m)} - \mathbf{Z}_{m,i}\boldsymbol{\gamma}_m \\ \mu_{y^*_{s,i}}^{(m)} - \mathbf{X}_{s,i}\boldsymbol{\beta}_s \\ \mu_{w^*_{s,i}}^{(m)} - \mathbf{Z}_{s,i}\boldsymbol{\gamma}_s \\ \mu_{y^*_{f,i}}^{(m)} - \mathbf{X}_{f,i}\boldsymbol{\beta}_f \\ \mu_{w^*_{f,i}}^{(m)} - \mathbf{Z}_{f,i}\boldsymbol{\gamma}_f \end{pmatrix}',$$

where $\sigma^{2(m)}_{ji} = \text{Cov}(y_{m,i}, \ldots, w_{f,i} | \boldsymbol{\theta}^{(m)}_j, \boldsymbol{\Omega}^{(m)}_j, \mathbf{y})$, $\mu^{(m)}_{y^*_{k,i}} = E[y^*_{k,i} | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Omega}^{(m)}_k, \mathbf{y}]$ $k = s, f, m$, $\mu^{(m)}_{w^*_{k,i}} = E[w^*_{k,i} | \boldsymbol{\theta}^{(m)}, \boldsymbol{\Omega}^{(m)}_k, \mathbf{y}]$ $k = s, f, m$. The elements in $\mathbf{Q}_{ji}$ associated to equations not solved by entities in regime j must be set equal to zero.

The Gibbs Sampler

Gibbs sampling (Casella and George 1992) is necessary to simulate the nonobserved information present in the matrices $\mathbf{Q}_{ji}$. The sampler requires the distribution of each $y^*_{k,i}$ and $w^*_{k,i}$ conditional on the values of the rest of the dependent variables in the corresponding regime. It is well known that these distributions are univariate normal when the unconditional multivariate distribution is normal. Let the means and variances of the conditional distributions at the $m + 1$ iteration be $\mu^{(m)}_{y^*_{k,i}|(-y^*_{k,i})}$, $\sigma^{2(m)}_{y^*_k|(-y^*_k)}$, $\mu^{(m)}_{w^*_{k,i}|(-w^*_{k,i})}$, and $\sigma^{2(m)}_{w^*_k|(-w^*_k)}$, respectively, where $_{|(-y^*_{k,i})}$ indicates conditionality on the values of all the other dependent variables (apart from $y^*_{k,i}$) being present in the regime at which entity i belongs.

Simulations for $y^*_{k,i}$ must be done conditional on its corresponding observed information $y_{k,i}$. The observed counterpart of $y^*_{k,i}$ is dichotomous with $y^*_{k,i}$ being positive if $y_{k,i}$ equals one and nonpositive if $y_{k,i}$ equals zero. Accordingly, we simulate $y^*_{k,i}$ from a normal distribution with mean $\mu^{(m)}_{y^*_{k,i}|(-y^*_{k,i})}$ and variance $\sigma^{2(m)}_{y^*_k|(-y^*_k)}$ truncated below at zero if $y_{k,i}$ equals one and truncated above at zero if $y_{k,i}$ equals zero. The observed counterparts of variables $w^*_{k,i}$ are categorical ordered and defined by equation 3. Correspondingly, we simulate $w^*_{k,i}$ from a normal distribution with mean $\mu^{(m)}_{w^*_{k,i}|(-w^*_{k,i})}$, and variance $\sigma^{2(m)}_{w^*_k|(-w^*_k)}$ truncated above at $\alpha_{k,t+1}$ and truncated below at $\alpha_{k,t}$ when $w_{k,i}$ equals $l_{k,t}$ ($k = s, f, m$; $t = 1, \ldots, r$).

A complete set of starting values $y^{*(0)}_{k,i}$ and $w^{*(0)}_{k,i}$ is required to initiate the Gibbs sampler. We use $y^{*(0)}_{k,i} = 0\ \forall k, i$ and $w^{*(0)}_{k,i} = w_{k,i}$. The simulation is then repeated iteratively until completing sequences $y^{*(1)}_{k,i}, \ldots, y^{*(K(m))}_{k,i}$ and $w^{*(1)}_{k,i}, \ldots, w^{*(K(m))}_{k,i}$, where $K^{(m)}$ is a number large enough to ensure convergence. Wei and Tanner (1990) recommend starting with a small $K^{(1)}$ and progressively increasing $K^{(m)}$ as m increases. Then eliminate a number k_{burn} of simulations from the beginning of the sequence. The remaining simulations in the sequence are used to estimate the terms $\sigma^{2(m)}_{ji}$, $\mu^{(m)}_{y^*_{k,i}}$, and $\mu^{(m)}_{w^*_{k,i}}$ in $\mathbf{Q}_{ji}$.

M-Step

Following Meng and Rubin (1993), it is advisable to replace the M-step by two conditional M-steps. The first conditional M-step maximizes $E[\ell^c(\boldsymbol{\theta}, \boldsymbol{\Omega} | \mathbf{y})]$

with respect to the elements in $\boldsymbol{\theta}$ conditional on $\boldsymbol{\theta}^{(m)}$ and $\boldsymbol{\Omega}^{(m)}$. After a little matrix calculus, it is easy to see that the maximizer in this first conditional maximization can be written as a generalized least squares estimator

$$\boldsymbol{\theta}^{(m+1)} = \left[\boldsymbol{X}_d' \left[\sum_j \left(\tilde{\boldsymbol{\Omega}}_j^{-1} \otimes \boldsymbol{I}^j \right) \right] \boldsymbol{X}_d \right]^{-1} \boldsymbol{X}_d' \left[\sum_j \left(\tilde{\boldsymbol{\Omega}}_j^{-1} \otimes \boldsymbol{I}^j \right) \right] \boldsymbol{\mu}_{y*}^{(m)},$$

where $\mathbf{I}^j$ is a $N \times N$ diagonal matrix with $\mathbf{I}_{ii}^j = 1$ if entity i belongs to regime j and $\mathbf{I}_{ii}^j = 0$ otherwise. The 6×6 matrix $\tilde{\boldsymbol{\Omega}}_j^{-1}$ contains the elements of $\boldsymbol{\Omega}_j^{-1}$ in the positions corresponding to the equations solved in regime j, while the remaining elements must be set equal to zero. The block diagonal matrix $\mathbf{X}_d$ is defined as

$$\mathbf{X}_d = \begin{bmatrix} \mathbf{X}_m & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_m & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{Z}_f \end{bmatrix}, \text{where } \mathbf{Z}_{k,i} = 0 \text{ if } y_{k,i} = 0; \boldsymbol{\mu}_{y*}^{(m)} \left(\boldsymbol{\mu}_{y_m^*}^{(m)} \ \boldsymbol{\mu}_{z_m^*}^{(m)} \ldots \boldsymbol{\mu}_{z_f^*}^{(m)} \right)^T,$$

$$\boldsymbol{\mu}_{y_k^*}^{(m)} = \left(\boldsymbol{\mu}_{y_{k,1}^*}^{(m)} \ \ldots \ \boldsymbol{\mu}_{y_{k,i}^*}^{(m)} \ \ldots \ \boldsymbol{\mu}_{y_{k,N}^*}^{(m)} \right)^T, \boldsymbol{\mu}_{z_k^*}^{(m)} = \left(\boldsymbol{\mu}_{z_{k,1}^*}^{(m)} \ \ldots \ \boldsymbol{\mu}_{z_{k,i}^*}^{(m)} \ \ldots \ \mu_{z_{k,N}^*}^{(m)} \right)^T \text{ and } \boldsymbol{\mu}_{z_{k,i}^*}^{(m)} = 0 \text{ if } y_{k,i} = 0.$$

The second conditional M-step estimates $\boldsymbol{\Omega}^{(m+1)}$ by maximizing $E[\ell^c(\boldsymbol{\theta}, \boldsymbol{\Omega} | \mathbf{y})]$ with respect to the elements in $\boldsymbol{\Omega}$ conditional on $\boldsymbol{\theta}^{(m+1)}$ and $\boldsymbol{\Omega}^{(m)}$. No closed form for $\boldsymbol{\Omega}^{(m+1)}$ exists; thus numerical optimization techniques must be used at this stage. Thresholds $\alpha_{k,3} < \cdots < \alpha_{k,r}$ are not present in the complete information likelihood function; therefore they cannot be obtained by first order condition or by numerical optimization. We proceed the following way to estimate $\alpha_{k,t}$. First, at every round of the Gibbs sampler at iteration m, keep the minimum value of every sequence obtained when simulating the observations $w_{k,i} = l_{k,t}$; this produces a set of $K^{(m)} - k_{burn}$ values. Second, keep the maximum value of every sequence obtained when simulating the observations $w_{k,i} = l_{k,t-1}$. Third, calculate the medians of the two sets obtained in the preceding two steps. Finally, take the average between the two medians, which produces a consistent estimator of $\alpha_{k,t}$. The E-step and M-step are then repeated until convergence is attained.

Appendix B. The Information Matrix

Louis's identity (Louis 1982) was used in this study to obtain a Monte Carlo estimation of the information matrix

$$\mathbf{I}(\boldsymbol{\theta};\mathbf{y}) = -\mathbf{H}^c(\boldsymbol{\theta};\mathbf{x}) - E\left[\mathbf{S}^c(\boldsymbol{\theta};\mathbf{x})\mathbf{S}^c(\boldsymbol{\theta};\mathbf{x})'\right] + E\left[\mathbf{S}^c(\boldsymbol{\theta};\mathbf{x})\right]E\left[\mathbf{S}^c(\boldsymbol{\theta};\mathbf{x})'\right]$$

where $\mathbf{H}^c(\boldsymbol{\theta};\mathbf{x}) = \dfrac{\partial^2 \ell^c(\boldsymbol{\theta};\mathbf{x})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ and $\mathbf{S}^c(\boldsymbol{\theta};\mathbf{x}) = \dfrac{\partial \ell^c(\boldsymbol{\theta};\mathbf{x})}{\partial\boldsymbol{\theta}}$ are the complete information Hessian and score vector, respectively. All of the expectations are estimated at the final MCEM estimators. Monte Carlo estimates of the complete information Hessian and score vectors can be used to estimate the information matrix (Ibrahim, Chen, and Lipsitz 2001).

Since thresholds $\alpha_{k,t}$ are not present in the complete information maximum likelihood, their standard errors cannot be obtained from the information matrix presented above. Following Albert and Chib (1993), we consider that estimates of $\alpha_{k,t}$ are uniformly distributed between the two medians calculated in the third step above in appendix A when estimating $\alpha_{k,t}$. Thus standard errors for our estimates of $\alpha_{k,t}$ were calculated as the square roots of the variances of those distributions.