Room for Discretion?
Biased Decision-Making in International Financial Institutions

Valentin F. Lang (Heidelberg University)

Andrea F. Presbitero (International Monetary Fund and MoFiR)

Abstract: We exploit the degree of discretion embedded in the World Bank-IMF Debt Sustainability Framework (DSF) to understand the decision-making process of international financial institutions. The unique, internal dataset we use covers the universe of debt sustainability analyses conducted between December 2006 and January 2015 for low-income countries. These data allow us to identify cases where the risk rating implied by the application of the DSF's mechanical rules was overridden to assign a different official rating. Our results show that both political interests and bureaucratic incentives influence the decision to intervene in the mechanical decision-making process. Countries that are politically aligned with the institutions' major shareholders are more likely to receive an improved rating; especially in election years and when the mechanical assessment is not clear-cut. These results suggest that the *room for discretion* international financial institutions have can be a channel for informal governance and a source of biased decision-making.

JEL codes: F34, F53, H63, H68

Keywords: International organizations; Political economy; IMF; World Bank; Debt sustainability

Valentin F. Lang, Heidelberg University; email: valentin.lang@awi.uni-heidelberg.de. Andrea F. Presbitero, International Monetary Fund and MoFiR; email: apresbitero@imf.org. We thank two anonymous referees, Luca Bandiera, Andrew Berg, Tito Cordella, Axel Dreher, Andreas Fuchs, Christopher Humphrey, Christopher Kilby, Sarah Langlotz, Gerard Padro i Miquel (the editor), Juan Pradelli, Katharina Richert, Eyal Rubinson, Tim Willems, several IMF and World Bank staff, and participants at the conference on the Political Economy of International Organizations, January 12-14, 2017, at the European Public Choice Society conference, April 19-22, 2017, the Development Economics and Policy conference, June 1-2, 2017, and at seminars at the IMF and at Heidelberg University for useful comments and discussions. We also thank numerous staff at the IMF for assistance in accessing the data on the debt sustainability analyses. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

1. Introduction

A considerable body of literature suggests that the major international financial institutions (IFIs) are not the independent, technocratic organizations many expect them to be. Instead of making decisions exclusively based on objective and economic criteria, their behavior also reflects the particular interests of individual actors, which appear to bias the organizations' decision-making in their favor.[1] Focusing on the two most powerful IFIs, the International Monetary Fund (IMF or Fund) and the World Bank, a first strand of literature shows that the political interests of major shareholders play a significant role in their lending decisions. Countries that are politically aligned with or important for the United States and other "G5" governments[2] have privileged access to financial assistance from the Fund and the Bank (Barro and Lee, 2005; Dreher et al., 2015, 2009a, 2009b; Dreher and Jensen, 2007; Kilby, 2013a, 2009; Reynaud and Vauday, 2009; Stone, 2008; Thacker, 1999). A second strand of literature identifies bureaucratic incentives as a specific source of bias in the IMF and World Bank decision-making processes. According to this view, which follows standard public choice models and approaches underlining the importance of organizational culture, staff in international organizations aim to increase their budget, power, prestige and independence. As a result, the decisions IFIs make are not always economically optimal and technocratic, but can also reflect the bureaucracy's particular interests and beliefs (Barnett and Finnemore, 2004; Chwieroth, 2013; Copelovitch, 2010; Nelson, 2014; Stone, 2008; Vaubel, 2006, 1996, 1986).

In this paper we exploit a specific feature of the design of the Debt Sustainability Framework (DSF) for low-income countries (LICs), developed jointly by the World Bank and the IMF to assess debt sustainability, as an ideal set-up to understand the decision-making of IFIs. We use a unique dataset on the application of the DSF that allows us to reconstruct internal decision-making processes and to identify decisions that deviated from the mechanical application of formal rules. In particular, the DSF assigns a rating (low risk, moderate risk, high risk) for the risk of debt distress to a given country according to the projected evolution of its debt levels over a 20-year period, with respect to policy-dependent debt thresholds. While the assessment of the risk of debt distress is based on a mechanical,

[1] For a recent survey of this literature see Dreher and Lang (2016).

[2] In addition to the United States, this group of powerful shareholders of the Bank and the Fund comprises the United Kingdom, France, Germany and Japan.

model-based rule, the final rating may involve the use of judgment by staff. The DSF explicitly allows staff to override the mechanical rating if they consider that country-specific circumstances justify this choice. In practice, in a number of cases this *room for discretion* translates into deviations of the final risk ratings from the ones obtained by mechanically applying the rule. Our data allow us to identify the cases in which such overrides of the mechanical rating took place. In the empirical analysis we examine the determinants of these overrides.

Specifically, we test for the presence of political and bureaucratic biases by looking at whether the country's political alignment with the organizations' major shareholders and the staff's desire to maintain its existing risk rating are related to overrides of the mechanical risk rating. In addition, we also investigate the role that other macroeconomic variables play in determining the actual risk rating, with the aim of shedding light on the design of the DSF, whose simplicity—a result of the need to make it accessible to a wider set of stakeholders—could result in a limited capacity to take into account important macroeconomic developments in the mechanical risk rating.

In the context of the literature on the political economy of international organizations (IOs) our empirical approach allows us to explicitly test the influential model of "informal governance" (Stone, 2013, 2011, 2008). This model, which has been initially developed for the IMF, posits that the influence of powerful states on international organizations primarily runs through informal channels.[3] While formal rules and a relatively autonomous bureaucracy regulate the IOs' day-to-day operations, powerful states retain influence through informal practices that allow them to intervene in formal processes when urgent strategic interests are at stake.[4] The model furthermore suggests that such informal influence must only be selectively exercised to avoid undermining the IO's legitimacy, which in turn is an important reason for powerful states to act through IOs in the first place. Transferred to our setting, the *room for discretion* embedded in the DSF is a potential channel for such informal political influence. Furthermore, it gives bureaucrats the opportunity to influence decision-making according to the bureaucracy's preferences. It would be consistent with this model's predictions if

[3] See Stone's (2011) book and the 2013 special issue of the Review of International Organizations (volume 8, issue 2) on "Informal Governance in International Organizations." See also Copelovitch (2010) for a related model, which posits "common agency" by the largest shareholders and the bureaucracy.

[4] In the words of Stone (2013: 124) "all international organizations operate to some degree at variance with their formal rules. The formal rules—standard operating procedures, voting rules, organizational chains of command, written policies—provide stable and predictable policy outputs. Derogations from these standard procedures are made to safeguard the interests of powerful states."

political interests and bureaucratic incentives were reflected in the use of judgment and if, in addition, the political bias were more significant a) when overruling is less clearly opposed to the formal rules, and b) when political interests are particularly strong. We test the former hypothesis by allowing for heterogeneity of the political interests effect depending on how clear-cut the mechanical rating is, and the latter by examining whether the political bias is stronger in election years of the rated country.

In addition to the explicit test of the informal governance model in a particularly suitable setting, the main contribution of our approach is threefold. First, we can directly examine the extent to which political economy variables explain interference in technocratic rules instead of only comparing how they relate to differences in outcomes. The deviations from the mechanical rule can be interpreted as a direct measure of the Bank and Fund's discretion in assessing the borrowing capacity of a country. In contrast to most of the previous literature, we can thus shed light on the internal decision-making processes that lead to the outcomes that IFIs produce.

Second, in contrast to many studies that analyze IMF and World Bank lending, the focus on risk ratings minimizes selection bias. Different countries might have different levels of demand for Fund and Bank resources. The fact, for instance, that countries that disagree with major shareholders on issues of foreign policy are also less likely to have an IMF program could be due to the lack of willingness to engage with the Fund, rather than to unfavorable treatment from the Fund and Bank. Since the DSF is a standard toolkit of IMF surveillance that is regularly applied in *all* low-income countries and all countries have the same interest in obtaining the most favorable risk rating under the DSF (as it would assure better borrowing terms, see below), there are no differences on the demand-side in this setting.

Third, while the analysis of the DSF is particularly well suited to study the decision-making process of international financial institutions, examining potential biases in DSF risk ratings is also a relevant policy question *per se*. The debt sustainability analyses—and the final risk ratings—provide a unique and relevant source of information for a variety of stakeholders on the sustainability of the fiscal stance in LICs, which are typically not covered by the major sovereign credit agencies.[5] The risk

[5] Our study thus also contributes to the literature on the determinants of credit ratings more broadly. Recent contributions have examined bias in sovereign credit ratings assigned by private rating agencies (Bartels and Weder di Mauro, 2013; Fuchs and Gehring, 2017).

ratings have financial and macroeconomic consequences, as they can directly influence the size and the terms of borrowing of low-income countries.[6]

Since most LICs do not have regular access to international capital markets, development finance is a major source of their external financing. The DSF risk ratings determine the terms under which countries receive financial assistance from multilateral institutions. For instance, the World Bank's International Development Agency (IDA) reduces its allocation to countries with weak ratings by up to a fifth and makes the loans-grants mix conditional on the risk rating.[7] In addition, the World Bank uses the DSF risk ratings to design non-concessional borrowing limits for LICs. To test whether the World Bank adheres to these rules in practice, we regress a country's volume of lending received from the World Bank (as a fraction of GDP) on its DSF risk rating and a basic set of control variables. We find that "medium risk" and "high risk" countries indeed receive significantly less lending; on average and holding other factors constant, "high risk" countries receive between 17 and 35 percent less financing than "low risk" countries (see Table A1 in the Appendix).[8] In the same spirit, regional development banks and bilateral aid agencies base their grant and lending decisions on DSF ratings. For instance, according to the so-called "Lagarde rule" (from when the current IMF Managing Director was Minister of Finance in France), the Agence Française de Développement only lends to countries at low risk of debt distress and continues to lend for a year in case a country is downgraded to moderate risk, while it does not lend at all to high risk countries. Finally, risk ratings affect debt conditionality under IMF-supported programs. Countries at moderate or high risk of debt distress

[6] We also contribute to the literature that goes beyond the analysis of IMF/World Bank lending activities, where the existing evidence on biasedness is much scarcer. Dreher et al. (2008), find that IMF inflation forecasts are systematically biased and favor countries that are politically close to the United States. Fratzscher and Reynaud (2011) find that countries with more political influence in the IMF and in the UN receive more favorable Public Information Notices of Article IV consultations from the Fund. Other than that there is little systematic evidence as to whether political and bureaucratic interests bias Fund and Bank decision-making in areas that are – at least *a priori* – less political and more technocratic in nature.

[7] Being classified as high risk is associated with 100 percent grants, medium risk with 50 percent grants and 50 percent loans, while low risk is associated with 100 percent loans and zero grants. Grants come with a 20 percent reduction in available resources to minimize moral hazard. See: http://ida.worldbank.org/financing/debt-sustainability-grants. The countries' preference for higher volumes of World Bank lending appears to dominate the relative allocation between grants and loans. One reason could be related to a substantial time discount rate that politicians attach to such financing due to political cycles and political incentives. Loans are highly concessional and come with very long maturities and grace periods. As of April 2017, the regular IDA loan has a 38-year maturity and a 6-year grace period. In this case, a politician could prefer borrowing 100 percent rather than receiving 80 percent in grants, and ignore the repayment problem, as it will be effective only after 6 years.

[8] These regressions also show that the negative effect of the official DSF rating on World Bank lending holds when controlling for the mechanical rating, suggesting that potential biases in the deviations from the mechanical rating would directly affect lending.

have different kinds of debt limits, while program conditionality in countries at low risk of debt distress normally does not include limits on public external borrowing (IMF, 2015).

In recent years, however, an increasing number of LICs have started gaining market access, mostly through syndicated bank loans and Eurobond issuances (Presbitero et al. 2016). For these economies, the DSF risk ratings are also a critical source of information for market participants and can affect the availability and the terms of commercial lending. A descriptive look at our data illustrates this. Out of 14 low-income countries that have issued sovereign bonds since 2014, ten were classified at a low risk of debt distress, only four had a moderate risk (Cameroon, Cote d'Ivoire, Ghana and Honduras), and no country with a high risk of debt distress has been able to issue. The average spread at issuance has been about 140 basis points higher for countries at moderate risk than for countries at low risk of debt distress, even taking into account the size and maturity of the bonds. Moreover, comparing countries with similar *Institutional Investor* country risk ratings but different DSF risk ratings shows that a worse DSF risk rating is indeed reflected in higher sovereign bond spreads. For instance, in 2014 and 2015, Ghana had a similar country risk rating as Mongolia, Senegal and Zambia, but it was the only country classified at moderate risk of debt distress and faced an average premium ranging between 110 and 228 basis points compared to the other three countries at low risk of debt distress.[9]

A similar picture emerges when considering cross-border bank lending. Over the 2007-2015 period the financial markets platform *Dealogic* records 803 syndicated loans to countries with an outstanding DSF risk rating, 59.4 percent of which went to countries with a low risk, 34.1 percent to countries with a moderate risk and only 6.5 percent to countries with a high risk of debt distress. This distribution is even more skewed towards low risk countries when considering loans to the private sector, three quarters of which have had a company in a low risk country as borrower. If we look at dynamics over time and focus on countries that have seen a worsening of their debt rating, we can see that countries that have been regular borrowers (Mali, Mongolia, Mozambique, Niger, and Sri Lanka, with at least one loan before and after a debt sustainability analysis) have on average experienced a reduction in

[9] More precisely, the country risk ratings by *Institutional Investor* range between 0 and 100 with 100 indicating the least likelihood of default (see http://www.institutionalinvestor.com/Research-and-Rankings.html). We compare Ghana—at moderate risk—with Mongolia, Senegal and Zambia—at low risk—over the period 2014-2015. In that period, the four countries had, on average, a very similar *Institutional Investor* country credit rating, but faced different sovereign bond spreads. In particular, the credit rating was 33.6 in Ghana, 33.2 in Mongolia, 33.3 in Senegal and 35.3 in Zambia, while the average spreads were 670 bps in Ghana, 556 in Mongolia, 441 in Senegal and 559 in Zambia.

the number of syndicated loans when considering a window of one year before and one year after the publication of the risk rating.[10]

Our results indicate that there is evidence for the presence of systematic biases. First, the debt ratings for countries that are politically aligned with the Bank's and the Fund's major shareholders are more often overridden (and improved) as compared to the result of the strict application of the DSF's mechanical model. The effect is stronger when the mechanical assessment is less clear-cut, suggesting that biasedness is more likely when there is more *room for discretion*. Also, the effect is mainly driven by the years in which an election took place in the rated country, suggesting that allies of the major shareholders are particularly likely to be treated favorably when their governments particularly care about positive assessments.[11] Second, bureaucratic incentives also explain the use of judgment in the DSF. In a large number of cases staff make use of the *room for discretion* in order to avoid deteriorations of the debt sustainability rating relative to the previous assessment. This is consistent with the expectation that the bureaucracy has an incentive to "keep" the ratings assigned in previous analyses as this is more likely to be in line with the established in-house view and avoids potential confrontations with senior staff. Finally, we find that some macroeconomic variables are associated with the decision to use judgment, suggesting that the DSF is "too simple" and Bank and Fund staff have to rely on additional information that is not part of the DSF to assess debt sustainability. We show that these results are unlikely to be driven by endogeneity, unobserved country-specific heterogeneity, and we rule out the possibility that other channels and alternative factors could explain the override of the risk rating. We also corroborate our statistical results with the help of anecdotal evidence that we gathered in interviews with involved Bank and Fund staff.

While we believe that our results document the presence of systemic biases, they are silent about the trade-off between rules versus discretion. We point to some of the potential costs of the discretionary bias, but we do not assess its benefits. For instance, overridden ratings could reduce the rate of false alarms and better reflect *ex-post* debt sustainability. Because endogeneity prevents us from such an

[10] The relatively small size of cross-border lending to LICs and the fact that we cannot disentangle demand from supply effects would call for caution when interpreting these numbers.

[11] This result is consistent with recent findings by Kersting and Kilby (2016, p. 153), who show that World Bank lending accelerates for allies of the United States when domestic elections approach, and interpret this as evidence for the World Bank's engagement in "global electioneering that serves U.S. foreign policy interests."

assessment—as future debt dynamics critically depend on actual risk ratings—we have to leave the analysis of the potential benefits of the *room for discretion* for future research.

2. The Debt Sustainability Framework and the risk ratings

The DSF was introduced in 2005 to guide the borrowing decisions of LICs in a way that matches their financing needs with their current and prospective repayment ability, taking into account each country's specific circumstances. Under the framework, debt sustainability analyses (DSAs) are conducted regularly with the aim of assessing the risk of external debt distress and providing policy recommendations that limit the risk of debt distress through prudent borrowing and lending strategies.[12] All DSAs must be prepared jointly by the World Bank and the IMF and submitted to their respective executive boards. Staff of the two institutions coordinate closely in preparing the DSAs from the early stages to the final approval. They are supposed to agree on the key macroeconomic projections and assumptions on new borrowing, with World Bank staff generally taking the lead on growth prospects and IMF staff focusing on medium-term macroeconomic projections. As a general rule, one DSA should be produced at least once every year (not necessarily at a given point in time for all countries), as it is the basis to determine the IDA credit-grant allocation. Exceptions to this time schedule may happen under specific circumstances, such as a request for IMF financing that involves exceptional access (IMF 2013). To empirically check whether this regular annual schedule is adhered to in practice, we look at the number of months between two consecutive DSAs for a given country. In only 19 cases two DSAs are separated by less than six months—possibly because the DSA has been triggered by a specific event. As DSAs appear to generally follow a regular, pre-determined schedule we can reliably rule out selection bias resulting from selection into receiving a rating.[13]

The DSA assigns a risk of debt distress (low risk, medium risk, high risk, or in debt distress) to a country depending on the evolution of five debt indicators compared to some policy-dependent thresholds, assuming that breaches of these thresholds would signal the presence of debt

[12] Since we recognize that the terminology may appear confusing, we would like to point out that "although the terms 'DSF' and 'DSA' are sometimes used interchangeably, they are in fact distinct: the DSF is the framework within which a DSA is produced for a particular country" (IMF 2013, p. 6).

[13] To minimize concerns about selection bias, we show that our results are robust to limiting the sample to the DSAs that were conducted between 6 and 18 months after the previous one and thus strictly followed the pre-determined schedule (see Section 4.4).

vulnerabilities. Specifically, the mechanical rule for assigning debt risk ratings is based on the projections of five key macroeconomic indicators for the subsequent 20 years: the present value (PV) of external debt as a percentage of GDP, exports, and revenue as well as the external debt service as a percentage of exports and revenue. For each of these five variables there is one baseline scenario and eight stress test projections with alternative macroeconomic assumptions for the 20-year period. The projected values are compared to the thresholds, that depend on the country's quality of policies and institutions, as measured by a three-step version of the Country Policy and Institutional Assessment (CPIA)—an index of institutional quality and macroeconomic stability produced by the World Bank.[14] According to the mechanical rule, the risk rating 'moderate risk' is assigned if one of the stress test projections exceeds the corresponding threshold, while the rating 'high risk' is assigned if this is the case for one of the baseline scenarios (IMF, 2013). In sum, these data allow us to reconstruct the rating suggested by the mechanical rule and to compare it to the rating actually assigned.[15]

The actual rating may differ from the mechanical one because of the use of judgment by IMF and World Bank staff. One of the distinctive features of the DSF is to explicitly allowing for judgment as a source of flexibility to ensure that ratings are not excessively affected by short term macroeconomic fluctuations and that they also take into account country-specific characteristics that are not reflected in the macroeconomic projections (IMF, 2009). In particular, the guidance notes of the LIC DSF state that "*[a]lthough the indicative thresholds play a fundamental role in the determination of the risk rating, they should not be interpreted mechanistically. The assessment of risk needs to strike a balance between paying due attention to debt levels rising toward or above thresholds and using judgment. Thus, a marginal or temporary breach of a threshold may not necessarily imply a significant vulnerability. Conversely, a near breach should not be dismissed without careful consideration*" (IMF, 2013). More precisely, the assessment done by the country team can deviate from the rule in the presence of minor and temporary breaches of the debt

[14] These policy-dependent thresholds are based on the estimation of a set of simple probit models on a large sample of low and middle income countries over the 1970-2007 period—one for each debt indicator—where the probability of debt distress is a function of the debt indicator, the CPIA indicator as a measure of policies and institutional quality, and GDP growth as a proxy for economic shocks. See IMF and World Bank (2012) for additional details.

[15] For an overview of the LIC DSF, see http://www.imf.org/external/np/exr/facts/jdsf.htm. For a more detailed discussion on the underlying method and its application, see IMF (2013) and Berg et al. (2014). After its introduction, the DSF has been reviewed three times (2006, 2009 and 2012), but its main structure has remained the same. The two most important changes introduced by the 2012 review have been: 1) the revision of two of the five debt thresholds (debt over exports and debt service over exports), to take into account remittances in the denominator of the ratio for countries highly dependent on remittances, and 2) the introduction of the "probability approach", which uses country-specific information to complement the standard debt sustainability assessment in borderline cases.

thresholds, while protracted (and large) breaches are more worrisome. Also, staff should pay attention to the pace of debt accumulation and give the right weight to some of the stress scenarios, which—being standardized—in some circumstances may not be very realistic. Moreover, careful consideration should be given to the country's ability to repay external debt that is not captured in the framework. For instance, the availability of a large pool of international reserves could counteract some negative indication coming from debt service ratios.

3. Data and Method

3.1 Main variables and descriptive evidence

Our analysis is based on a unique dataset covering *all* 377 debt sustainability analyses undertaken in low-income countries between December 2006 and January 2015. The dataset includes, for each DSA, the five debt ratios in the year of the DSA and their projections for the 20-year period ahead, as well as all projections for the key macroeconomic variables that are used to determine the evolution of the country's level of debt flows and stocks. These projections are the ones used by IMF and World Bank staff in the macroeconomic framework produced by the country team at the time when the DSA was produced. Some information on the projections over the medium term (but not the yearly data for the entire 20-year period) for the key variables are publicly available in the country report and in the associated DSA write-up.

We are interested in the determinants of the decision to make use of the discretion embedded in the DSF. In Table 1 we take a first, descriptive look at our data on the DSAs by comparing mechanical and actual ratings. First, the table shows that in the majority of cases the actual rating is equivalent to the mechanical rating. Of the 367 DSAs for which we have all necessary data,[16] 272 end up with the rating that the mechanical rule suggests (~74%). Making use of the *room for discretion* in the DSF is the exception rather than the rule. Second, it becomes evident that the vast majority of deviations are improvements of the risk rating (i.e., a lower risk) with respect to the ones mechanically determined. Almost all of these deviations are one-notch improvements (i.e., from high risk to medium risk or

[16] The last DSF revision introduced remittances-augmented debt thresholds and projections to capture enhanced repayment capacity of countries receiving large remittances (see footnote 15). Because of incomplete data for the ten DSAs done in 2014 and 2015 that use these augmented values, we exclude them from the analysis.

from medium risk to low risk); only one country received a low risk rating when the projections suggested assigning high risk. The actual rating is worse than the mechanical rating in only nine cases.[17] This first piece of evidence would already suggest the possibility that the almost unidirectional use of judgment is consistent with the presence of staff incentives to provide a good rating of the country they work on, as well as with a positive bias that could be explained by political pressures to obtain an improvement in the risk rating.

Against this backdrop, for the remainder of our analysis we code a binary dependent variable that indicates whether the final DSF rating was more favorable than what the mechanical rating suggested and examine the determinants of such rating improvements. While we consider multiple potential determinants, we are primarily interested in two hypotheses that emerge from the political economy literature on international organizations.

First, we test the hypothesis that the use of judgment in debt sustainability ratings is biased in favor of countries that are politically affiliated with the Bank's and the Fund's major shareholders. For such a bias to be present, shareholders do not have to interfere directly with the decision-making processes at the staff level where DSAs are produced. The mechanisms may be more subtle. Preemptive obedience to the views of the shareholders' delegates (Executive Directors) may play a role as all DSAs have to be approved by the Board of Directors. In our background research at the institutions' headquarters in Washington, staff suggested that both the Fund and the Bank have certain "narratives" for most countries and treat some countries as role model for others. Internally, DSAs are expected to fit these preexisting "stories" for individual countries. Such narratives and role models can originate at the political level and trickle down via senior management to the staff level.[18] While the exact mechanisms remain speculative, it is clear that such favorable treatment of the major shareholders' political allies would constitute a bias that is not reconcilable with a correct application of the DSF that is solely based on economic criteria.

To measure political proximity we use data on voting behavior in the United Nations General Assembly (UNGA). In our baseline, we rely on estimates of the distances between the "ideal point" of

[17] We looked at the justification for the ratings of these nine DSAs in detail and found that the analysts in all cases referred to country-specific vulnerabilities. Examples include the 2014 DSA for Afghanistan, which concluded that risk was high because of "significant vulnerabilities" including the country's heavy reliance on donor grants, and the 2012 DSA for Burkina Faso, which set the risk rating from low to moderate because of vulnerabilities related to the country's gold exports. Because of the small sample size, we do not test for potential biases in these nine decisions.

[18] Source: interviews with World Bank and IMF officials (November 2016).

rated countries and the United States (Bailey et al., 2017).[19] Similar to Vreeland and Dreher (2014) we code a country as a "US-friend" if it is in the lowest quintile of the distribution of the ideal point distance (meaning that it is among the closest friends of the United States), as we expect variation at the upper tail of the political alignment distribution to be more likely to make a difference than variation around the mean. As our study focuses on LICs, most of which score low on political alignment with the United States, we believe that minor differences in political alignment between the United States and the rated country are uninformative. By contrast, we expect benefits to materialize only for "close friends," which are the ones that are significantly different from the average LIC. We test the robustness of this assumption using alternative cutoff values as well as the mean distance to ideal points of the United States, the United Kingdom, Germany, France and Japan, the so-called G5. In Table 2 we descriptively compare US-friends to other countries and find that the former are indeed more likely to benefit from the use of judgment in DSAs (42 vs. 29 percent)—a first indication of bias but nothing close to rigorous econometric evidence.

Second, we consider the role of bureaucratic incentives. Existing literature and our background research at the World Bank and the IMF suggests that there often is a strong disincentive for staff to increase a country's risk of debt distress relative to its previous rating. First, as a worse rating implies that lending from the World Bank must be reduced, the "pressure to lend" (Kilby, 2009, 2000) incentivizes the Bank bureaucracy to "show that its customers are solvent."[20] Second, an increase in the DSF risk rating constitutes a break with past analyses and suggests that previous projections had to be adjusted. Staff involved in producing new DSAs confirmed that they have an incentive to agree with the conclusions drawn in previous analyses, as any change requires a stronger justification vis-à-vis management than maintaining the *status quo*. Moreover, some countries function as important role models for the institutions and staff may fear that increasing the debt risk ratings for such countries may undermine the consistency and credibility of the institutions' policy analysis and advice. To statistically test whether staff use the *room for discretion* in the DSF to avoid increasing countries' debt risk ratings, we code a binary variable (*Risk Increased*) that indicates that the mechanical rating is worse than the actual rating assigned in the previous DSA (i.e. increasing from low risk to

[19] In contrast to other voting affinity scores such as S-scores these ideal points are better at removing noise by using information on changes in the UNGA's agenda (Bailey et al., 2017). The variables are defined as ideal point distances, such that a smaller value indicates greater voting affinity.

[20] Source: Interview with World Bank official (November 2016)

moderate/high risk, or from moderate risk to high risk). We look at these cases in Table 3. It shows that out of the 80 DSAs in which assigning the mechanical rating would have meant increasing the country's risk rating relative to its previous DSA, in 64 DSAs (80 percent) the mechanical rating was overridden to keep the rating the same as in the previous DSA.[21] The descriptive evidence thus strongly supports the hypothesis that the IMF and the World Bank make use of judgment in order to make DSF risk ratings more persistent at lower risk levels than the strict application of the formal framework would suggest.[22]

While this descriptive evidence suggests the presence of both a bureaucratic and a political bias, we can neither rule out that this bias is driven by confounding factors, nor can we illuminate any underlying channels. In order to do so, we turn to a more rigorous econometric analysis.

3.2 Econometric Model

We run binary response regressions of the variable indicating an improvement of the DSF risk rating employing a conditional logit estimator that controls for unobserved time-invariant heterogeneity across regions or countries via fixed effects. We also include year fixed effects to absorb the impact of the global business cycle and other global factors that can affect all LICs equally. We naturally exclude all observations whose ratings cannot be further improved and also control for the mechanical rating as the conditional propensity to override might be different for adjustments from high to medium as compared to adjustments from medium risk to low risk. Formally we estimate:

$$\Pr\left(A_{itn} < M_{itn} \middle| \alpha_{r/i}, \delta_t, X_{itn}, M_{itn}\right)$$

where subscripts *i, t, r,* and *n* denote countries, years, regions and DSAs respectively. *M* and *A* denote the mechanical and the actual risk rating, α and δ are full sets of region (or country) and year fixed effects and X_{itn} is a vector of explanatory observable variables that are either country-year-specific or

[21] When contrasting this with the mirror case, we find that of the 24 cases in which the mechanical rating suggested a lower risk than before, eight (33 percent) were overridden to leave the rating unchanged. As this fraction is positive yet substantially lower than in the opposite case, this evidence suggests that there is both a conservative bias to keep the previous rating and a bias of avoiding to increase risk ratings.

[22] We know the subsequent rating for 50 of the 64 DSAs that were overruled when the mechanical rating suggested a deterioration. 21 of these 50 reverted mechanically back to a better rating in the subsequent DSA while for the remaining 29 the mechanical rating again suggested a higher risk. Note also that in the data there is no regular pattern of how many times a worse mechanical rating is tolerated before a worse official rating is assigned.

DSA-specific.[23] The explanatory variables of interest in X_{itn} are the country-year specific measure of political proximity (*US friend*$_{it}$) and the DSA-specific variable indicating that the mechanical rule suggests assigning a worse rating relative to the previous DSA (*Mechanical Rating: Risk Increased*$_{itn}$). The effect of the latter variable can be estimated with country fixed effects, since it naturally varies significantly over time. However, given that political alignment with the US is highly persistent over time, we choose region fixed effects over country fixed effects in the baseline regression with the *US friend* variable. Moreover, since regional spillovers are a key element of the spread of financial (and sovereign debt) crises, controlling for regional effects serves to capture these region-wide unobserved effects. We expect the effect of political proximity to the US in a ten-year panel to be driven primarily by cross-country variation, which would be absorbed by country fixed effects.[24]

To exploit within-country variation over time and to address endogeneity concerns in these regressions, we include interaction effects with time-varying variables. First, as theoretical considerations suggest that the political bias is stronger when the mechanical rating is less clear-cut, we allow the effect to be heterogeneous depending on the DSA-specific, time varying number of threshold breaches. Second, we interact the measure of political alignment with major shareholders with a variable (*Election*$_{it}$) indicating years in which an election took place in the rated country, as we expect governments to be particularly interested in favorable assessments in election years.[25] In addition to exploiting within-country variation over time—allowing us to employ country fixed effects and thus significantly mitigating the concern that there could be unobserved country-specific, time-invariant factors affecting the likelihood of obtaining better ratings—this approach has a second methodological advantage. Under the assumption that the timing of elections is exogenously predetermined, the interaction can be interpreted in a causal way. Even if political proximity were endogenous to the decision to improve the risk rating after conditioning on all covariates, the main effect of political proximity would be biased, but the interaction would still have a causal

[23] Note that we use the subscript *itn* because in some, rare cases there are multiple DSAs for the same country within one year. Countries *i* are nested in regions *r*.

[24] Another advantage of using region fixed effects is that with country fixed effects we lose almost half the sample because several countries never receive an improved rating. However, we include country fixed effects in additional specifications where we interact the US friend variable with a dummy for the presence of elections.

[25] In an alternative specification we use a variable indicating that an election took place in the six months following the publication of a DSA. Arguably, political interests in good ratings are not only generally strong in election years but especially strong *in the run-up* to an election.

interpretation (Nizalova and Murtazashvili 2016). We consider endogenous election timing in our setting to be highly unlikely, as we are not aware of any instances in which an election was rescheduled because of an upcoming DSF rating.[26] In addition, even if there were unobserved variables correlated with both elections and overrides of mechanical DSF ratings—due to anticipation effects—such endogeneity would only bias the interaction effect if it was conditional on political proximity to the United States.[27] In sum, we believe that the identifying assumption necessary for a causal interpretation of the interaction is plausible.

In addition to the explanatory variables of interest, we augment X_{itn} with the following covariates.[28] We first add control variables that derive naturally from the DSF's method. The mechanical rating is a function only of threshold breaches and the CPIA. We test whether these two variables also affect the decision to override the mechanical rating. According to the DSF guidance note (IMF 2013), staff are explicitly supposed to take the number of threshold breaches into account: "[l]arge, protracted breaches are more worrisome than small, temporary ones. Breaches of multiple thresholds suggest greater vulnerabilities than a single breach." We thus expect more overrides in DSAs with fewer threshold breaches. Specifically, we use the number of threshold breaches in the baseline and stress test projections, scale these two figures by the number of possible breaches of the baseline test (5 x 20 = 100) and the stress tests (8 x 5 x 20 = 800) and add the two fractions up to generate the variable *Breaches*.[29]

Second, we control for the continuous CPIA index, as the DSF's method emphasizes the importance of policies and institutions by making the thresholds contingent on the CPIA, and by emphasizing that the "quality of policies and institutions has a large influence on the macroeconomic return of public investment" (IMF 2013: 24). We expect that the continuous version of the CPIA provides additional information that is not captured by the discrete, three-step version of the index used for the

[26] In similar settings, Faye and Niehaus (2012) as well as Kersting and Kilby (2016) do not find evidence for endogenous election timing with regard to aid inflows. This suggests that endogenous election timing with regard to upcoming DSF ratings is also highly unlikely.

[27] Instead of assuming $E(\text{election} \cdot \varepsilon) = 0$ one would then assume $E(\text{election} \cdot \varepsilon) = E(\text{election} \cdot \varepsilon \mid \text{US friend})$. For econometric details on this point, see Bun and Harrison (2014) and Appendix S4 in Dreher et al. (2016).

[28] See Appendix A2 for sources, descriptions and descriptive statistics of all variables.

[29] Fund staff suggested that the stress test based on the historical scenario (which fixes some key macroeconomic variables at their historical averages to compute the 20-years ahead projections) is, in practice, often considered to be implausible given that many countries are expected to be on a structurally different growth path compared to the mid-1990s and/or early-2000s. To address this concern, we also ran regressions in which we ignore this stress test. Our inferences are not affected.

determination of the debt thresholds. Moreover, Fund and World Bank staff may consider the CPIA as a proxy for "other country-specific considerations" that should be taken into account when gauging whether a few breaches warrant downgrades in risk ratings.

We add other macroeconomic fundamentals to test if they also influence the use of judgment and to rule out that they confound our findings. We focus on variables that the literature identifies as typical determinants of sovereign risk ratings (Hill et al., 2010) and control for the logarithm of *GDP per capita*, *GDP growth*, and the logarithm of *population*, as larger countries are less sensitive to economic shocks than small ones. In additional regressions we also control for the level of total public debt, scaled by gross national income, and for the current account balance over GDP. Then, as the guidance note advises to take into account whether the country has large foreign exchange reserves when applying judgment (IMF 2013), we test if the level of international *reserves* can contribute to explaining the overrides. In the same spirit, we control for *natural resource rents* to account for the idea that debt sustainability could hinge on countries' future revenues from natural resources, information which cannot easily be incorporated in the DSF and which therefore could require the use of judgment.[30] In a robustness test, we also look at the effect of *IMF growth projections*, as optimistic growth scenarios could not only lower debt ratios but could also motivate an overrule of the mechanical risk rating.[31]

As the DSF's method explicitly leaves room for the application of judgment, it would be in accordance with an unbiased application of the DSF's method if any of these control variables were related to the decision to override the mechanical result of the DSF. In general, however, we expect the omitted variable bias that is driven by macroeconomic fundamentals to be low, given that the projections used to produce the DSAs already incorporate most of this information. Any correlations between the macroeconomic fundamentals and the decision to manually adjust the risk rating would show that staff rely on information that is not incorporated in the mechanical projection models. This would suggest that the DSF's method is "too simple" and would support a refinement of the econometric underpinnings of the DSF (Berg et al., 2014).

[30] Note that we lag all macroeconomic explanatory variables by one year to make sure that they were observed at the time the DSA was produced.

[31] All of these variables are taken from the World Bank's World Development Indicators (WDI), with the exceptions of growth projections which are taken from the IMF World Economic Outlook (WEO) database (different vintages).

4. Results

4.1 Macroeconomic Determinants

Before turning to our main variables of interest, in Table 4 we examine the effect of macroeconomic fundamentals and variables that are internal to the DSAs. Some of these variables could be correlated with overrides if the DSF is applied as intended, given that the number of threshold breaches and "country-specific considerations" are supposed to be taken into account. These regressions also help us to identify a suitable set of control variables.

As column 1 shows, the share of breached thresholds is a statistically significant predictor of the decision to override the risk rating. As one would expect, the higher this share, the less likely it is that the rating is improved. The continuous version of the CPIA also appears to influence this decision. Given that the three-step version of the CPIA plays an important role in setting the relevant thresholds in the DSF, but forces the framework to ignore some relevant information contained in the continuous measure of the CPIA, it is not surprising that Bank and Fund staff also rely on the additional information conveyed by the continuous version when deciding on whether to improve the rating. We also find that DSAs in which the mechanical rule suggests that a "high risk" rating be assigned are more likely to be improved, when the other variables are kept fixed. All three DSA-specific variables are statistically significant and enter with consistent signs across various specifications in Table 4 (and Table 5).

In column 2 we add GDP per capita, GDP growth and population as explanatory variables. As one would expect, GDP per capita appears as a strong, robust and statistically significant predictor of the outcome variable and enters with the expected sign in Table 4 (and Table 5). The statistical evidence for the effect of the growth rate is less robust (but in Table 5 has the expected, positive sign in the regressions in which it appears as a statistically significant predictor of overrides). Country size as measured by logged population is positively correlated with rating improvement.[32] While this could be interpreted as indicating a bias in favor of geopolitically important countries, large countries are more likely to be economically diversified and thus less vulnerable to external shocks. Moreover, larger countries could rely relatively more on local currency borrowing (Berensmann et al., 2015)

[32] In auxiliary regressions, which are available upon request, we find that the latter effect is driven by the largest countries in the sample.

therefore being less exposed to the risk of external default. In our view, this finding may thus be economically justified and is not necessarily an indication of bias. In column 3 we include two additional macroeconomic variables that are typical determinants of sovereign credit risk. The current account balance (as a share of GDP) and the external-debt-to-GNI ratio neither enter with statistically significant coefficients, nor do they affect the coefficients on the other explanatory variables. In column 4 we add the country's amount of reserves (scaled by country debt) and its rents from natural resources (as a share of GDP). Countries with larger amounts of reserves and natural resources have the option to use these assets in order to mitigate the risk of debt distress. In particular, reserves are highly liquid and can be used to service external debt, making some breaches of the debt service indicators less relevant for the DSF risk rating. Although the signs on the two variables are, as could be expected, positive, there is no statistically significant empirical evidence that suggests that this plays a role for the use of judgment in DSAs.

We conclude that in addition to the variables that are internal to each DSA (number of breached thresholds, CPIA, level of the mechanical rating), GDP per capita, and to a lesser extent growth and country size, help predicting improvements of the risk rating. Other macroeconomic fundamentals do not seem to matter. As the projections and the threshold breaches on which the ratings are based include a large amount of country-specific economic information, the finding that other macroeconomic fundamentals are not correlated with overrides of the mechanical rating is not surprising. Instead, it shows that the method underlying the DSF is successful in incorporating a large amount of country-specific information that is typically considered to be informative for assessments of debt sustainability. The fact that at least the CPIA and GDP per capita robustly predict rating overrides indicate that there is room for further improving the DSF. Nevertheless, the results presented so far suggest that, to a significant extent, the DSF as well as the *room of discretion* embedded in it are employed as intended.

4.2 Biases

In Table 5 we augment our baseline model with a set of variables that, if statistically associated with overrides of the mechanical DSF ratings, indicate the presence of bias. In these regressions our baseline specification builds on the covariates included in column 2 of Table 4, given that these

variables allow us to keep the entire sample, while the additional controls are not robustly related to the dependent variable.[33]

Initially, we examine the role played by politics. In column 1 we add the *US friend* indicator[34] and find that countries that are politically affiliated with the United States are more likely to benefit from the use of judgment in DSAs. In column 2 we allow this effect to be heterogeneous depending on the number of threshold breaches and find that the effect is much stronger when the projections breach only relatively few thresholds. Figure 1 illustrates this result. While for US friends the conditional likelihood for receiving an improved rating is more than 80 percent when less than 2 percent of all projections breach thresholds, the likelihood for other countries is around 40 percent. The more breaches that are projected the smaller the difference between these two groups of countries. The difference vanishes at around six percent of breached thresholds.[35] These findings suggest that risk ratings are more likely to reflect geopolitical preferences the less clear-cut the mechanical rating is.

While this distinct pattern gives us confidence that our *US friend* indicator does not pick up an unobserved confounder, we aim to get closer to a causal interpretation of the effect in column 3. We interact *US friend* with a variable that indicates whether the DSA was published in an election year in the rated country. The results of specification 3 show that the US friends' benefit from the use of judgment is driven by the election years in the sample. Elections in other countries are not related to rating improvements. As discussed above, even if we allow the assumption of conditional exogeneity of political proximity to be violated—and thus cannot infer that US friends are on average favorably treated—we can still conclude that elections cause differences in the favorable use of judgment only for the group of countries classified as US friends.

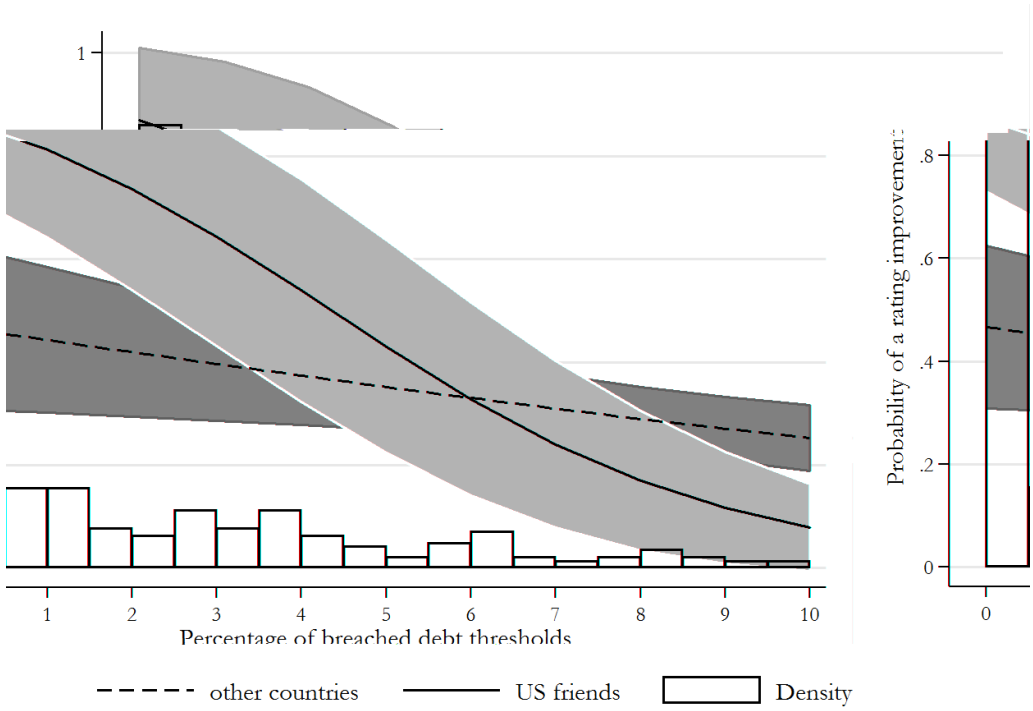
---

[33] All other variables we consider suffer from missing data and thus result in list-wise deletion of the respective observations. Given that our baseline regressions include 263 observations, we argue that the bias introduced by potentially non-random list-wise deletion of dozens of observations is likely to be more severe than omitting covariates that are found to not be robustly correlated with the outcome. In a robustness tests (Table 7), we show that the results do not depend on this choice.

[34] Note that these variables are lagged by two years instead of one as this avoids losing 37 observations and because most votes in the UNGA take place in November and December. As the correlation of the variables indicating ideal point distances from the G5 for t-1 and t-2 is $r = 0.94$ this is unproblematic and in the robustness section we show that our results do not depend on this choice.

[35] As the underlying density function of the percentage of breached thresholds shows, there are very few observations for this percentage being above ten. The results for these values are thus of no economic significance. The reason why the graph for US friends approaches zero for values close to and greater than ten is that there are no observations of US friends that receive an upgrading with this number of breaches. The model therefore correctly predicts a zero probability of overriding.

Figure 1: Visualizing the effect of political alignment with the United States



Notes: The line chart shows the estimated probability (and the 90 percent confidence intervals) of a risk rating improvement for countries classified as "US friends" and for other countries, depending on the percentage of debt thresholds breached in the DSA. The line chart is based on the results of a logistic regression with region dummies, rather than the ones of the conditional logistic regression, as reported in Table 5 column 2. The results of these two regressions are almost identical except that the standard errors are smaller when the conditional logit estimator is used. This is why the graph understates the precision of the estimation. All control variables are set at their sample means. The bar chart shows the density function of the breached debt thresholds, at different 0.5 percent bins.

Next, we consider the effect of bureaucratic incentives (column 4). As the descriptive evidence already suggested, the probability that the mechanical rating is overruled increases significantly whenever the projections indicate a deterioration of debt sustainability. The *room for discretion* is used to avoid assigning higher risk ratings relative to the previous assessment. The effect is economically substantial; if debt risk increases, the probability that the rating is overruled increases by 45 percentage points.[36]

[36] Effect size estimated via a linear probability model with country fixed effects based on column 9.

In column 5-7 we jointly test our two main hypotheses by replicating specifications 1-3 and adding the *Risk Increased* indicator. While in column 5 the coefficient on the *US friend* indicator is imprecisely estimated, all the other results, which exploit time variation, are robust to this modification. We conclude that – consistent with the "informal governance" model – bureaucratic and political biases are two distinct patterns.

Finally, in column 8-10 we replicate our main results (columns 3, 4, and 7) exploiting the within country variation and saturating the model with country fixed effects. While this specification is quite demanding, especially in a short panel, and substantially reduces the sample size, our results are still valid, so that we can confidently rule out that our main finding are driven by unobserved time-invariant country-specific characteristics.[37]

claims of BIS reporting banks from the United States.[38] Our results do not show any statistically significant evidence that bank exposure is related to the use of judgment to improve DSF risk ratings (column 1), even though the sign of the coefficients on the bank exposure variable and its interaction with the number of breaches are consistent with the presence of US economic and financial interests (column 2).

In the subsequent regressions we examine whether the political bias is in fact a cultural bias. In the literature on the political economy of sovereign credit ratings, Fuchs and Gehring (2017) have recently shown that the cultural distance between the home country of ratings agencies and the rated country affects the assessment of credit risk. Rating agencies perceive countries that are more distant in cultural and linguistic terms to be at higher risk than other comparable countries. Given that the World Bank's and the IMF's headquarters are both based in the United States, their staff are familiar with the United States' culture and language. If familiarity in cultural and linguistic terms indeed introduces familiarity bias that leads to a more favorable perception of rated countries and if culturally close countries are also politically closer to the United States we might misattribute a cultural familiarity bias to a political bias. In columns 3 and 4 we thus add a variable measuring a country's distance-adjusted ethno-linguistic fractionalization (DELF) relative to the United States to the regressions (Kolo, 2012).[39] Again, we find no statistically significant evidence for such a bias and rule out that the political bias we identify is in fact driven by cultural and ethno-linguistic familiarity.

In the literature on the political economy of the IMF it is well-established that countries with political ties to the United States are more likely to receive loans from the IMF. The finding that "US-friends" are more likely to benefit from the use of judgment in DSAs might thus be an artefact of the fact that the same countries are also more likely to be under an IMF arrangement. There are several reasons that explain why countries that receive IMF resources might be treated differently in DSAs. In line with the reasoning in Dreher et al. (2008) and Reynaud and Fratzscher (2011), IMF staff might have an interest in 'defensive rating', causing them to cast countries that receive IMF resources in a good light. At the same time, however, they might want to show that the country is at risk of debt distress to justify the existence (and a possible extension) of the program. When we include a variable indicating

[38] We run the same regressions with data on bank exposure for banks from the European Union and find the same results.

[39] While this variable is the best measure we have for the cultural distance between the rated country and the IMF and World Bank, we admit that it is a rough proxy as their staff are very international. These results should thus be taken with a grain of salt.

the presence of an IMF program in the year of the DSA, we initially find that such countries are less likely to receive improved ratings. As countries under IMF programs, however, see their mechanical DSF ratings deteriorate less frequently than countries not receiving Fund resources (39 vs. 61 percent), we control for *Risk Increased* and then find this effect to disappear. These findings are consistent with the following interpretation: the Fund adjusts the debt sustainability outlook for program countries less frequently than for non-program countries, since this would mean contradicting the projections on which the design of their programs was based. This is why overrides, on average, are also less frequent. However, if the projections are still adjusted under IMF programs, then judgment is used to avoid a downgrade: In fact, out of the 32 DSAs which suggested a deterioration of debt sustainability while the country was under an IMF program, 28 (87.5 percent) were overruled in order to keep the old rating. In sum, we interpret this result as further evidence for the bureaucratic bias and—given that the coefficient on the US-friend variable is not affected—we can rule out the possibility that the political bias is due to the higher propensity to participate in IMF programs.

4.4 Robustness

In Tables 7-9 we present a set of robustness exercises. We start by considering a number of additional factors that could motivate the use of judgment and possibly bias our estimates of the importance of political interests and bureaucratic incentives (Table 7).

First, in columns 1-3 we show that adding the larger vector of covariates (from Table 4, column 4) does not affect the results, notwithstanding a significant reduction in sample size. In addition to the current account balance, the debt-to-GNI ratio, the country's level of international reserves and its rents from natural resources, in these regressions we also consider the inflows of official development assistance from the US to rule out that the favorable treatment of US allies is due to the larger amount of US aid that such countries typically receive. We also add the interaction *CPIA x Breaches* to control for a potential non-linearity of the CPIA's effect depending on the number of threshold breaches.

Second, we include information on whether the rated country participated in the Heavily Indebted Poor Countries (HIPC) initiative and on whether it issued sovereign bonds at the time of the assessment. The former could signal a propensity to override the mechanical debt rating for reasons that have nothing to do with a political bias, but are instead related to a benign view of countries which received debt relief, for which the IMF and the World Bank may have incentive to show that

post-HIPC debt levels are sustainable. With regards to the latter, the fact that countries that issue sovereign bonds have easier access to external financing could also influence the assessment of debt sustainability and, in particular, the use of judgment. In fact, the specific structure of sovereign bonds—with bullet repayment at maturity—translates into sharp but temporary spikes in debt service, which could result in breaches of the relevant debt thresholds. Nevertheless, staff do not generally consider those breaches as relevant for debt sustainability and judgment is often used to overrule the mechanical rating. Those two variables, however, are not statistically significant and, more importantly, the coefficients on our variables of interest are not affected (columns 4-6).[40]

An alternative source of bias could arise from the fact that country teams can easily improve the risk rating by modifying the macroeconomic framework to incorporate overly optimistic growth projections. Not controlling for IMF growth projections may bias the estimate of our key variables, especially because IMF growth projections could themselves be politically biased (Dreher et al. 2008). To rule out that a bias in growth forecasts drives our findings, we add the difference between IMF growth projections and the actual growth rate to the baseline regressions.[41] Again, our results are robust to this modification (columns 7-9). One may also argue that ratings could improve because of a political bias in the CPIA. However, this is unlikely to be an issue in this context, given that the CPIA scores are computed by the World Bank for its aid allocation and—differently from growth projections—they enter the DSA as an exogenous input, over which the country team producing the DSA does not have any control. In any case, if there were a political bias in the IMF growth projections and in the CPIA, these biases would work against our results, so that our estimates can be safely considered a lower bound.

Finally, in columns 10-12 we remove all control variables except the DSA-specific variables and region and year fixed effects to compare how much the covariates affect the coefficient of interest. Following the method proposed by Altonji et al. (2005), we can then calculate "selection ratios" that indicate how much of the effect is explained by selection on observables. Comparing the "full" specifications (columns 1-3) to the specifications that only includes the core covariates (columns 10-12, same sample)

[40] However, in the first specification, which almost exclusively exploits cross-country variation, the standard error increases, resulting in the coefficient on *US friend* marginally missing statistical significance at the ten percent level ($p = .105$).

[41] We can compute this difference only over a one year horizon, because of data availability (extending the horizon would implicate a dramatic reduction in sample size, as we would have to focus only on "old" DSAs). However, we argue that if there is a strategic manipulation of growth projections, this should be already reflected in the first year.

we find the selection ratios to be 5.4 for the simple effect of the US friend indicator, 4.3 for the level effect (and 30.2 for the interaction effect) when allowing for non-linearity in threshold breaches, and 10.1 for the effect in election years. This suggests that selection on unobservables in the first two specifications would have to be more than four times larger than selection on observables for the effects to be zero. Given that we already control for a large amount of potential confounders, we think that it is unlikely that selection on unobservables drives this entire effect. The coefficient on the variable indicating US-friends with domestic elections is only marginally affected by the inclusion of covariates; the finding that selection on unobservables would have to be more than ten times as large as selection on observables to wash away this effect is consistent with our assumption that this interaction term is conditionally exogenous.

In sum, these findings are in line with our expectation that the endogeneity bias potentially induced by omitted variables is not substantial in this setting. After all, the DSF projections already include most of the relevant country-specific information, explaining why many macroeconomic fundamentals we observe do not substantially affect the decision to overrule. Furthermore, both the fact that the non-linearity of the effect with respect to the number of breaches behaves as theory predicts, and the evidence that the effect is driven by election years—whose timing is exogenously predetermined—make us confident that our results are not due to endogeneity.

As discussed in the introduction, the timing of DSAs is also generally exogenous and follows an annual cycle. However, under specific circumstances, DSAs can be triggered by requests for exceptional IMF financing. While these cases are relatively rare in our sample, in Table 8 we test the robustness of our findings by excluding the DSAs not undertaken between six and 18 months after the previous one and thus potentially outside the regular, pre-determined schedule. Results are again confirmed (columns 1-3). In column 4, we recode the election dummy to consider exclusively those elections that took place within the six months following the publication of a DSA. We hypothesize that political interests in good ratings should be especially strong in the run-up to an election. It supports this hypothesis that with this more restrictive definition of the election dummy, we find an

even stronger positive association between the *US friend* indicator and the likelihood of obtaining an improved risk rating in election periods.[42]

Finally, Table 9 reports the results of additional robustness exercises that use alternative definitions of the variable indicating political proximity to the major shareholders of the Bank and the Fund. First, we set alternative cutoff values to be classified as a "US friend" (top 25 percent and top 15 percent instead of top 20 percent). Second, we take a one year lag instead of a two year lag, at the cost of reducing the sample size. Third, we consider the mean of ideal point distances to all G5 countries instead of only the United States. Our main findings are robust to all these modifications.

4.5 Extension: Testing Manipulation of Projections based on Density Discontinuity

The projections of debt levels underlying the DSF are based on a standardized econometric framework. Nevertheless, IMF and World Bank economists have several degrees of freedom as to how this framework is applied: minor changes in assumptions concerning, e.g., the expected growth rate, can substantially affect the number of threshold breaches. During our background research, staff thus suggested that biases could not only be present in decisions to override the mechanical rating, but also in the projections underlying the mechanical rating.

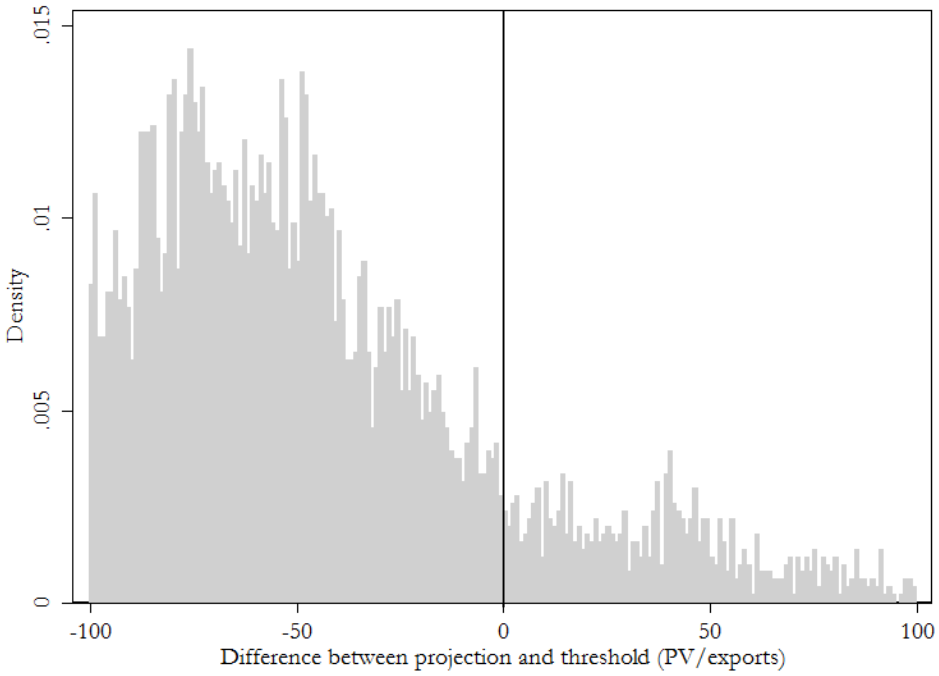
We argue that if such manipulation took place it would be reasonable to expect bunching of projections just below the relevant threshold. In other words, the density of a variable $\Delta = p - t$ measuring the difference between projections p and relevant thresholds t would be low for values just above zero and high for values just below zero.

To test this hypothesis we rely on manipulation testing in a regression discontinuity design – an idea introduced by McCrary (2008). We employ a nonparametric test for a discontinuity in the density of Δ at the threshold (in our case zero). Specifically, we use the manipulation test developed by Cattaneo et al. (2016a, 2016b), which is based on a local polynomial density estimator that does not require pre-binning the data (see also Calonico et al., 2014). This choice allows us to take a purely data-driven approach to estimating the density near the cutoff and to test hypotheses regarding the density's discontinuity.

[42] In this specification, countries not classified as US friends appear to have a slightly lower probability to benefit from improved risk ratings in the run-up to an election. This could be considered weak evidence for attempts to politically weaken incumbent governments that are not politically aligned with the United States (see Kersting and Kilby 2016).

We run this test for each of the five baseline projections (Table 10). In the sample of US-friends the hypothesis that the density is continuous at the cutoff can be rejected for two of the five projection sets at the 10 percent level and for one set at the 5 percent level. For two of these three sets of projections with potential discontinuities (*PV of debt over revenue* and *debt service over revenue*), however, the density estimator for values exceeding the threshold is based on very few observations, leading to potentially unreliable results. For the projections of *PV of debt over exports*, however, the test suggests a discontinuity of the variable's density at the relevant threshold. Interestingly, this indicator is the one that exceeds the threshold most often and seems to have the strongest informative content among the five, since it signaled the risk of debt distress in about 80 percent of all high risk cases. Figure 2 shows the histogram of Δ for *PV of debt over exports*, whose particular shape around zero is consistent with the hypothesis of bunching just below the relevant thresholds. Nevertheless, we find the evidence too weak to reliably exclude the possibility that this distribution of projections around the thresholds came about by chance. In sum, as we can only reject the null hypothesis at the 10 percent level for one of the five tests, we do not interpret this as definitive evidence for manipulation of projections. With reference to our main research question, however, we note that the size of the biases we find is, if anything, a lower bound as there is both anecdotal and weak statistical evidence that these biases also affect the estimation of projections and not only the decision to override the mechanical rating.

Figure 2: Testing Manipulation of Projections



Note: The graph is a histogram of a variable that measures the difference between the projected value of the debt indicator "present value of debt over exports" and the respectively relevant threshold. It visualizes the discontinuity of the density around 0 that the manipulation test (Table 10) reports.

5. Conclusion

Political interests and bureaucratic incentives influence the decision-making of international financial institutions. While our findings confirm previous research that has documented such biases, our approach and data allow us to reveal evidence on the specifics of how these biases operate. Our results suggest that the *room for discretion* embedded in the technocratic rules that aim to ensure objectivity enables political interests and bureaucratic incentives to influence (and bias) the decision-making of IFIs. We find that the influence of political interests is stronger when formal rules are less clear-cut. This evidence supports scholars claiming that "informal governance" and "unwritten rules" in international organizations enable powerful stakeholders to intervene in their decision-making (Kilby, 2013b; Koremenos, 2013; Stone, 2011).

In the particular empirical setting we consider, we find that the assessments under the World Bank's and the IMF's Debt Sustainability Framework for low-income countries—the only tool that provides

these countries with such debt risk ratings—are not free of bias. Both the geopolitical interests of the organizations' major shareholders and bureaucratic incentives appear to be reflected in the ratings. These countries, however, have a strong and legitimate interest in unbiased and objective assessments of their debt sustainability.

These biased ratings might misguide the lending decisions of various—public and private—creditors and could thus entail adverse economic effects. The use of discretion may also, however, reduce the number of false alarms. Assessing the actual presence and extent of positive and negative economic effects is complicated by a number of identification challenges. First and foremost, the fact that the rating *per se* affects future economic and financing conditions makes future debt sustainability endogenous. Absent such an assessment and a discussion of the potential benefits of the use of discretion, our results point to some potential costs of discretion, but do not provide a normative statement of the trade-off between rules and discretion. We leave this exercise for future research.

References

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. J. Polit. Econ. 113, 151–184.

Bailey, M.A., Strezhnev, A., Voeten, E., 2017. Estimating Dynamic State Preferences from United Nations Voting Data. J. Conflict Resolut. 61, 430–456.

Bank for International Settlements, 2016. Consolidated Banking Statistics.

Barnett, M., Finnemore, M., 2004. Rules for the World. Cornell University Press, Cornell.

Barro, R.J., Lee, J.W., 2005. IMF programs: Who is chosen and what are the effects? J. Monet. Econ. 52, 1245–1269.

Bartels, B., Weder di Mauro, B., 2013. A Rating Agency for Europe - A good idea?, Centre for Economic Policy Research Working Paper.

Beck, T., Clarke, G., Groff, A., Keefer, P., Walsh, P., 2001. New tools in comparative political economy:

The Database of Political Institutions. World Bank Econ. Rev. 15, 165–176.

Berensmann, K., Dafe, F., Volz, U., 2015. Developing local currency bond markets for long-term development financing in Sub-Saharan Africa. Oxford Rev. Econ. Policy 31, 350–378.

Berg, A., Berkes, E., Pattillo, C., Presbitero, A.F., Yakhshilikov, Y., 2014. Assessing Bias and Accuracy in the World Bank- IMF's Debt Sustainability Framework for Low-Income Countries. IMF Work. Pap. 14/48.

Bird, G., 1996. The International Monetary Fund and developing countries: A review of the evidence and policy options. Int. Organ. 50, 477–511.

Broz, J.L., 2005. Congressional politics of international financial resources. Am. J. Pol. Sci. 49, 479–496.

Broz, J.L., Hawes, M.B., 2006. Congressional Politics of Financing the International Monetary Fund. Int. Organ. 60, 367–399.

Bun, M., Harrison, T., 2014. OLS and IV estimation of regression models including endogenous interaction terms. UvA-Econometrics Work. Pap.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica 82, 2295–2326.

Cattaneo, M.D., Jansson, M., Ma, X., 2016a. rddensity: Manipulation Testing based on Density Discontinuity. Stata J. 1–18.

Cattaneo, M.D., Jansson, M., Ma, X., 2016b. Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. mimeo.

Chwieroth, J.M., 2013. "The silent revolution:" How the staff exercise informal governance over IMF lending. Rev. Int. Organ. 8, 265–290.

Copelovitch, M.S., 2010. Master or servant? common agency and the political economy of IMF lending. Int. Stud. Q. 54, 49–77.

Dreher, A., Eichenauer, V., Gehring, K., 2016. Geopolitics, Aid, and Growth. World Bank Econ. Rev.

Dreher, A., Jensen, N.M., 2007. Independent Actor or Agent? An Empirical Analysis of the Impact of U.S. Interests on International Monetary Fund Conditions. J. Law Econ. 50, 105–124.

Dreher, A., Lang, V.F., 2016. The political economy of international organizations. CESifo Work. Pap. 6077.

Dreher, A., Marchesi, S., Vreeland, J.R., 2008. The political economy of IMF forecasts. Public Choice

137, 145–171.

Dreher, A., Sturm, J.-E., Vreeland, J.R., 2015. Politics and IMF Conditionality. J. Conflict Resolut. 59, 120–148.

Dreher, A., Sturm, J.-E., Vreeland, J.R., 2009a. Global horse trading: IMF loans for votes in the United Nations Security Council. Eur. Econ. Rev. 53, 742–757.

Dreher, A., Sturm, J.-E., Vreeland, J.R., 2009b. Development aid and international politics: Does membership on the UN Security Council influence World Bank decisions? J. Dev. Econ. 88, 1–18.

Faye, M., Niehaus, P., 2012. Political Aid Cycles. Am. Econ. Rev. 102, 3516–3530.

Fratzscher, M., Reynaud, J., 2011. IMF surveillance and financial markets-A political economy analysis. Eur. J. Polit. Econ. 27, 405–422.

Fuchs, A., Gehring, K., 2017. The Home Bias in Sovereign Ratings. J. Eur. Econ. Assoc. 0, 1–38.

Gould, E.R., 2006. Money Talks: The International Monetary Fund, Conditionality, and Supplementary Financiers. Stanford University Press, Stanford.

Gould, E.R., 2003. Money Talks: Supplementary Financiers and International Monetary Fund Conditionality. Int. Organ. 57, 551–586.

Hill, P., Brooks, R., Faff, R., 2010. Variations in sovereign credit quality assessments across rating agencies. J. Bank. Financ. 34, 1327–1343.

IMF, 2016a. World Economic Outlook.

IMF, 2016b. Monitoring of Fund Arrangements (MONA) Database.

IMF, 2015. Staff Guidance Note on the Implementation of Public Debt Limits in Fund-Supported Programs, IMF Policy Papers. Washington, DC.

IMF, 2013. Staff Guidance Note on the Application of the Joint Bank-Fund Debt Sustainability Framework for Low-Income Countries, IMF Policy Papers. Washington, DC.

IMF, 2009. A Review of Some Aspects of the Low-Income Country Debt Sustainability Framework, IMF Policy Paper. Washington, DC.

IMF, World Bank, 2012. Revisiting the Debt Sustainability Framework for Low-Income Countries. Policy Paper, IMF and The World Bank, IMF and World Bank Policy Paper.

Kersting, E.K., Kilby, C., 2016. With a little help from my friends: Global electioneering and World Bank lending. J. Dev. Econ. 121, 153–165.

Kilby, C., 2013a. The political economy of project preparation: An empirical analysis of World Bank projects. J. Dev. Econ. 105, 211–225.

Kilby, C., 2013b. An Empirical Assessment of Informal Influence in the World Bank. Econ. Dev. Cult. Change 61, 431–464.

Kilby, C., 2009. The political economy of conditionality: An empirical analysis of World Bank loan disbursements. J. Dev. Econ. 89, 51–61.

Kilby, C., 2000. Supervision and performance: The case of World Bank projects. J. Dev. Econ. 62, 233–259.

Kolo, P., 2012. Measuring a New Aspect of Ethnicity – The Appropriate Diversity Index. Univ. Göttingen Discuss. Pap. 221.

Koremenos, B., 2013. What's left out and why? Informal provisions in formal international law. Rev. Int. Organ. 8, 137–162.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. J. Econom. 142, 698–714.

Nelson, S., 2014. Playing favorites: How shared beliefs shape the IMF's lending decisions. Int. Organ. 68, 297–328.

Nizalova, O., Murtazashvili, I., 2016. Exogenous Treatment and Endogenous Factors: Vanishing of Omitted Variable Bias on the Interaction Term. J. Econom. Methods 5, 71–77.

Nordhaus, W.D., 1975. The Political Business Cycle. Rev. Econ. Stud. 42, 169–190.

Oatley, T., Yackee, J., 2004. American Interests and IMF Lending. Int. Polit. 41, 415–429.

Presbitero, A.F., Ghura, D., Adedeji, O., Nije, L., 2016. Sovereign bonds in developing countries: Drivers of issuance and spreads. Rev. Dev. Financ. 6, 1–15.

Reynaud, J., Vauday, J., 2009. Geopolitics and international organizations: An empirical study on IMF facilities. J. Dev. Econ. 89, 139–162.

Stone, R.W., 2013. Informal governance in international organizations: Introduction to the special issue. Rev. Int. Organ. 8, 121–136.

Stone, R.W., 2011. Controlling Institutions: International Organizations and the Global Economy. Cambridge University Press, Cambridge.

Stone, R.W., 2008. The Scope of IMF Conditionality. Int. Organ. 62, 589–620.

Thacker, S.C., 1999. The High Politics of IMF Lending. World Polit. 52, 38–75.

USAID, 2016. U.S. Overseas Loans and Grants (Greenbook).

Vaubel, R., 2006. Principal-agent problems in international organizations. Rev. Int. Organ. 1, 125–138.

Vaubel, R., 1996. Bureaucracy at the IMF and the World Bank: A Comparison of the Evidence. World Econ. 19, 195–210.

Vaubel, R., 1986. A public choice approach to international organization. Public Choice 51, 39–57.

Vreeland, J.R., Dreher, A., 2014. The Political Economy of the United Nations Security Council. Cambridge University Press, Cambridge.

World Bank, 2016. World Development Indicators.

Tables

Table 1 – LIC-DSA mechanical and actual risk ratings

		Mechanical risk rating							
		Low		Moderate		High		*Total*	
Actual risk rating	Low	81	92.0%	47	30.7%	1	0.8%	*129*	*35.1%*
	Moderate	7	8.0%	104	68.0%	38	30.2%	*149*	*40.6%*
	High	0	0.0%	2	1.3%	87	69.0%	*89*	*24.3%*
	Total	*88*	*100.0%*	*153*	*100.0%*	*126*	*100.0%*	*367*	*100.0%*

Note: The table shows the absolute frequencies of mechanical and actual DSF risk ratings. The columns distinguish between the three categories of the mechanically produced ratings while the rows distinguish between the actual ratings countries have received. To the right of the absolute frequencies are relative frequencies that show the distribution of actual ratings given the mechanical rating.

Table 2 – Overruling and US friends

		US friend					
		No		Yes		*Total*	
Rating improved through overruling	No	158	71.2%	30	57.7%	*188*	*68.6%*
	Yes	64	28.8%	22	42.3%	*86*	*31.4%*
	Total	*222*	*100.0%*	*52*	*100.0%*	*274*	*100.0%*

Note: The table shows both absolute and relative frequencies of improved ratings given whether the country is classified as a US friend or not. Only DSAs with "moderate risk" and "high risk" mechanical ratings are considered as only these can be improved. Note that we lose five observations because of missing data on UNGA voting.

Table 3 – Overruling and avoiding downgrades

		Mechanical rating: risk increased					
		No		Yes		*Total*	
Mechanical rating adjusted downward	No	177	88.9%	16	20.0%	*193*	*69.2%*
	Yes	22	11.1%	64	80.0%	*86*	*30.8%*
	Total	*199*	*100.0%*	*80*	*100.0%*	*279*	*100.0%*

Note: The table shows both absolute and relative frequencies of improved ratings given whether the mechanical rating was worse than the country's previous DSF risk rating. Only "moderate risk" and "high risk" mechanical ratings are considered as only these can be improved.

Table 4 – DSA-specific variables and macroeconomic fundamentals

	(1)	(2)	(3)	(4)
Breaches	-0.116***	-0.102***	-0.105***	-0.105***
	(0.029)	(0.027)	(0.028)	(0.030)
CPIA	1.519***	1.232***	1.337***	1.647***
	(0.394)	(0.421)	(0.450)	(0.509)
Mechanical Rating: High	2.116***	2.222***	2.412***	2.488***
	(0.486)	(0.500)	(0.524)	(0.539)
GDP per capita (ln, t-1)		1.060***	1.178***	0.993***
		(0.297)	(0.322)	(0.360)
GDP growth (t-1)		-0.017	-0.032	-0.050
		(0.033)	(0.033)	(0.035)
Population (ln, t-1)		0.407***	0.461***	0.380**
		(0.136)	(0.152)	(0.170)
Debt/GNI (t-1)			0.004	0.003
			(0.008)	(0.010)
Current Account Balance (%GDP, t-1)			-1.044	-1.355
			(1.900)	(1.961)
Reserves/Debt (t-1)				0.004
				(0.005)
Natural Resources (%GDP, t-1)				0.023
				(0.015)
Observations	263	263	255	247
Pseudo R-squared	0.189	0.242	0.258	0.272

Note: dependent variable indicates an improved debt rating through overruling; conditional logistic regressions that control for region fixed effects; reported are the coefficient estimates of the regressions; standard errors, robust to clustering at the region level, in parentheses. All regressions include year fixed effects. Significance levels: * p<.10, ** p<.05, *** p<.01

	"#$!	"%$!	"&$!	"'$!	"($!	")$!	"*$!	"+$!	",$!	"#-$!
./!012345!"67%$!	0.421***!	2.166***!	-0.110!	!	0.280!	1.660***!	-0.005!	-2.905!	!	-4.011!
	(0.122)!	(0.386)!	(0.164)!	!	(0.377)!	(0.415)!	(0.443)!	(1.867)!	!	

!

	"#$!	"%$!	"&$!	"'$!	"($!	")$!	"*$!	"+$!
. /!0123 45!"67%$!	-^'%*!	#^+)(__!	-^'-,_!	%^-),__!	-^&%&__!	%^##-__!	-^%)&!	#^)(%__!
	"-^&#%$!	"-^)-,$!	"-^#,-$!	"-^',)$!	"-^-,%$!	"-^&+&$!	"-^&,#$!	"-^&#%$!
. /!0123 45!"67%$!!	!	7-^&()__!	!	7-^&*+__!	!	7-^&,&__!	!	7-^'&&__!
\!913:;<3=!	!	"-^-,,$!	!	"-^#%+$!	!	"-^###$!	!	"-^#))$!
. /!U:4E=!	7-^%)-!	#-^,#)!	!	!	!	!	!	!
"a K L H!67#$!	"#%^&&($!	",^*-+$!	!	!	!	!	!	!
. /!U:4E=!!	!	7#^),,!	!	!	!	!	!	!
\!913:;<3=!	!	"'^&&)$!	!	!	!	!	!	!
L>SR!6@!. /!	!	!	7-^-,)!	7-^)%*!	!	!	!	!
	!	!	"#^-(,$!	"#^)(&$!	!	!	!	!
L>SR!6@!. /!!	!	!	!	-^-+*!	!	!	!	!
\!913:;<3=!	!	!	!	"-^#&#$!	!	!	!	!
FA R!M1@C1:Y!	!	!	!	!	7#^-'#__!	7-^)+%__!	7-^%-,!	-^%,(!
	!	!	!	!	"-^&'+$!	"-^&-#$!	"-^(&+$!	"-^%*#$!
FA R!M1@C1:Y!!	!	!	!	!	!	7-^-)(!	!	7-^-))!
\!913:;<3=!	!	!	!	!	!	"-^-(*$!	!	"-^-')$!
A 3;<:42;:?!B:624C0!!	!	!	!	!	!	!	'^'-(__!	'^'#&__!
B2=E!F4;13:=35!	!	!	!	!	!	!	"-^'%)$!	"-^'#'$!
G@461@?=!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!
TU=31V:62@4=!	%(,!	%(,!	%(,!	%(,!	%(,!	%(,!	%(,!	%(,!
H=3P5@!B7=WP:135!	-^%&)!	-^%,'!	-^%&)!	-^%+,!	-^%(+!	-^&#)!	-^('&!	-^(*'!

X@630!53M345346!V:12:U?3!2452;:63=!:4!2YM1@V35!53U6!1:624C!6<1@PC<!@V311P??24CZ!;@45262@4:?!?@C2=62;!13C13==2@4=!6<:6!;@461@?!
0@1!13C2@4!02\35!3003;6=Z!13M@1635!:13!6<3!;@3002;2346!3=62Y:63=!@0!6<3!13C13==2@4=Z!=6:45:15!311@1=N!1@UP=6!6@!;?P=6312 4C!:6!6<3!
13C2@4!?3V3?N!24!M:1346<3=3=^!9:=3??43!;@461@?=!24;?P530!913:;<3=N!GHF I N! A 3;<:42;:?!B:624C0! J 2C<N! K L H!M31!;:M26:!"?4N!67#$N!
K L H!C1@O6<!"67#$N!H@MP?:62@4!"?4N!67#$^! I ??!13C13==2@4=!24;?P53![3:1!02\35!3003;6=^!/2C4202;:4;3!?3V3?=0!_!M`^#-N!__!M`^-(N!__!M`^-#!

!

!

!

!

!

!	"#$!	"%$!	"&$!	"'$!	"($!	")$!	"*$!	"+$!	",$!	"#-$!	"##$!	"#%$!
. /!012345!"67%$!	-^((,__!	#^++(__!	-^##*!	-^'#)!	%^#,'__!	7-^#(%!	-^&-%_!	#^,(#__!	7-^&%(!	-^)+)__!	%^')&__!	-^%''!
!	"-^-,*$!	"-^%(+$!	"-^%&%$!	"-^%(*$!	"-^(,&$!	"-^%,%$!	"-^#)%$!	"-^()-$!	"-^%&($!	"-^#(-$!	"-^#,,$!	"-^#*&$!
. /!012345!"67%$!8!U13:;<3=!	!	7-^&+-__!	!	!	7-^&+)__!	!	!	7-^&*+__!	!	!	7-^&,&__!	!
!	!	"-^#''$!	!	!	"-^#%-$!	!	!	"-^#&($!	!	!	"-^##-$!	!
. /!012345!"67%$!8!3?3;62@4!	!	!	%^&%-__!	!	!	%^'--__!	!	!	%^(,*__!	!	!	%^-,#__!
!	!	!	"-^*&-$!	!	!	"-^)(#$!	!	!	"-^+('$!	!	!	"-^*''$!
>?3;62@4!	!	!	7-^#,-!	!	!	7-^&,&!	!	!	7-^&**!	!	!	7-^#,,!
!	!	!	"-^'(+$!	!	!	"-^%'#$!	!	!	"-^%&)$!	!	!	"-^&&,$!
GP11346!I;;@P46!9:?:4;3!	7#^)+&!	7-^+%,!	7#^(&-!	!	!	!	!	!	!	!	!	!
"aKLHN!67#$!	"#^&+&$!	"#^+-,$!	"#^&)*$!	!	!	!	!	!	!	!	!	!
L3U6]KXF!"67#$!	-^--)!	-^--)_!	-^--)_!	!	!	!	!	!	!	!	!	!
!	"-^--'$!	"-^--'$!	"-^--&$!	!	!	!	!	!	!	!	!	!
B3=31V3=]L3U6!"67#$!	-^--)!	-^--(!	-^--)!	!	!	!	!	!	!	!	!	!
!	"-^--)$!	"-^--*$!	"-^--($!	!	!	!	!	!	!	!	!	!
X:6P1:?!B3=@P1;3=!	-^-#(!	-^--+!	-^-#%!	!	!	!	!	!	!	!	!	!
"aKLHN!67#$!	"-^-#)$!	"-^-#&$!	"-^-#*$!	!	!	!	!	!	!	!	!	!
./!:25!	7&^,&#!	7&^&+'!	7&^,&'!	!	!	!	!	!	!	!	!	!
"aKLHN!67#$!	"&^%+*$!	"&^%'*$!	"&^#)-$!	!	!	!	!	!	!	!	!	!
GHFI!8!913:;<3=!	-^#&%__!	-^#&-!	-^#&(__!	!	!	!	!	!	!	!	!	!
	"-^-))$!	"-^-+,$!	"-^-)+$!	!	!	!	!	!	!	!	!	!
JFHG!"LH$!	!	!	!	7-^(%)!	7-^)((!	7-^',%!	!	!	!	!	!	!
!	!	!	!	"-^*')$!	"-^*,'$!	"-^*-%$!	!	!	!	!	!	!
/@V3132C4!9@45=!	!	!	!	7-^,&#!	7-^(%'!	7#^#%#!	!	!	!	!	!	!
!	!	!	!	"-^+*%$!	"#^-'($!	"-^+#&$!	!	!	!	!	!	!
>11@1!24!FAR!K1@O6<!	!	!	!	!	!	!	-^--*!	-^-#%!	-^-#'!	!	!	!
R@13;:=6=!	!	!	!	!	!	!	"-^-&($!	"-^-'%$!	"-^-&*$!	!	!	!
I 55262@4:?! G@461@?=!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	L/I7 =M3;202;!	L/I7 =M3;202;!	L/I7 =M3;202;!
XPYU31!@0!TU=31V:62@4=!	%''!	%''!	%''!	%(,!	%(,!	%(,!	%(#!	%(#!	%(#!	%''!	%''!	%''!
H=3P5@!B7=WP:135!	-^&-#!	-^&&'!	-^&#*!	-^%''!	-^%,&!	-^%)(!	-^%'%!	-^%+*!	-^%)*!	-^#,+!	-^%(&!	-^%#)!

X@63D!53M345346!V:12:U?3!2452;:63=!:41?YM1@V35!53U6!1:624C!6<1@PC<!@V311P??4CZ!;@45262@4:?!?@C2=62;!13C13==2@4=!6<:6!;@461@?!0@1!13C2@4!02\35!3003;6=Z!13M@1635!:13!6<3!;@3002;2346!3=62Y:63=!@0!6<3!13C13==2@4=Z!=6:45:15!311@1=N!1@UP=6!6@!;?P=63124C!:6!6<3!13C2@4!?3V3?N!24!M:1346<3=3=^!9:=3??43!;@461@?=!24;?P53D!913:;<3=N!GHFIN!A3;<:42;:?!B:624CD!J2C<N!KLH!M31!;:M26:!"?4N!67#$N!KLH!C1@O6<!"67#$N!H@MP?:62@4!"?4N!67#$^!L/I7=M3;202;!;@461@?=!24;?P53D!913:;<3=N!GHFIN!A3;<:42;:?!B:624CD!J2C<^!I??!13C13==2@4=!24;?P53![3:1!02\35!3003;6=^!/2C4202;:4;3!?3V3?=D!_!M`^#-N!__!M`^-(N!___!M`^-#!

	@4?[!13CP?:1?[!62Y35!L/I=! "U36O334!)!:45!#+!Y@46<=!:0631!M13V2@P=$!			@4?[!3?3;62@4=!24!!)!Y@46<=!0@??@O24C!L/I!
!	"#$!	"%$!	"&$!	"'$!
./!012345!"67%$!	-^*#+__!	%^%,*__!	-^#),!	7-^-',!
	"-^#%+$!	"-^&)&$!	"-^#(-$!	"-^#(&$!
./!012345!"67%$!8!U13:;<3=!	ı	7-^&(*__!	ı	ı
	ı	"-^-)($!	ı	ı
./!012345!"67%$!8!3?3;62@4!	ı	ı	%^*#-__!	ı
	ı	ı	"-^)%#$!	ı
>?3;62@4!	ı	ı	7-^#'#!	ı
	ı	ı	"-^&%%$!	ı
./!012345!"67%$!\!>?3;62@4!	!	ı	ı	&^%,+__!
"24!)!Y@46<=!0@??@O24C!L/I$!	ı	ı	ı	"-^+,,$!
>?3;62@4!"24!)!Y@46<=!0@??@O24C!L/I$!	ı	ı	ı	7-^((,_!
	ı	ı	ı	"-^&%#$!
G@461@?=!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!
XPYU31!@0!TU=31V:62@4=!	%--!	%--!	%--!	%(,!
H=3P5@!B7=WP:135!	-^%*(!	-^&#(!	-^%,,!	-^%)'!

X@63D!53M345346!V:12:U?3!2452;:63=!:4!2YM1@V35!53U6!1:624C!6<1@PC<!@V311P?24CZ!;@45262@4:?!?@C2=62;!13C13==2@4=!6<:6!;@461@?!0@1!13C2@4!02\35!3003;6=Z!13M@1635!:13!6<3!;@3002;2346!3=62Y:63=!@0!6<3!13C13==2@4=Z!=6:45:15!311@1=N!1@UP=6!6@!;?P=63124C!:6!6<3!13C2@4!?3V3?=N!24!M:1346<3=3=^!9:=3??43!;@461@?=!24;?P53D!913:;<3=N!GHFIN!A3;<:42;:?!B:624CD!J2C<N!KLH!M31!;:M26:!"?4N!67#$N!KLH!C1@O6<!"67#$N!H@MP?:62@4!"?4N!67#$^!I??!13C13==2@4=!24;?P53![3:1!02\35!3003;6=^!/2C4202;:4;3!?3V3?=D!_!M`^#-N!__!M`^-(N!___!M`^-#!

!	"#$!	"%$!	"&$!	"'$!	"($!	")$!	"*$!	"+$!	",$!	"#-$!	"##$!	"#%$!
./!012345!"6@M%(a$!	-^'++_!	#^+%,__!	-^#-(!	!	!	!	!	!	!	!	!	!
!	"-^%(*$!	"-^(+%$!	"-^%))$!	!	!	!	!	!	!	!	!	!
./!012345!"6@M%(a$!\!	!	7-^&#+__!	!	!	!	!	!	!	!	!	!	!
913:;<3=!	!	"-^#''$!	!	!	!	!	!	!	!	!	!	!
./!012345!"6@M%(a$!\!	!	!	#^+*+__!	!	!	!	!	!	!	!	!	!
>?3;62@4!	!	!	"-^,-,$!	!	!	!	!	!	!	!	!	!
./!012345!"6@M#(a$!	!	!	!	-^(%%__!	%^&*,__!	7-^%#+!	!	!	!	!	!	!
!	!	!	!	"-^#('$!	"-^,*-$!	"-^&-*$!	!	!	!	!	!	!
./!012345!"6@M#(a$!\!	!	!	!	!	7-^&*-!	!	!	!	!	!	!	!
913:;<3=!	!	!	!	!	"-^%%,$!	!	!	!	!	!	!	!
./!012345!"6@M#(a$!\!	!	!	!	!	!	%^()-__!	!	!	!	!	!	!
>?3;62@4!	!	!	!	!	!	"-^,*%$!	!	!	!	!	!	!
./!012345!"67#$!	!	!	!	!	!	!	-^*)&__!	%^#&,__!	-^#%(!	!	!	!
!	!	!	!	!	!	!	"-^%,&$!	"-^(++$!	"-^&*#$!	!	!	!
./!012345!"67#$!8!	!	!	!	!	!	!	!	7-^&-%_!	!	!	!	!
913:;<3=!	!	!	!	!	!	!	!	"-^#)+$!	!	!	!	!
./!012345!"67#$!\!	!	!	!	!	!	!	!	!	%^-,#__!	!	!	!
>?3;62@4!	!	!	!	!	!	!	!	!	"-^++-$!	!	!	!
K(!012345!"67%$!	!	!	!	!	!	!	!	!	!	-^),+__!	%^(%#__!	-^%))!
!	!	!	!	!	!	!	!	!	!	"-^##($!	"-^'%($!	"-^#)&$!
K(!012345!"67%$!8!	!	!	!	!	!	!	!	!	!	!	7-^'-&__!	!
913:;<3=!	!	!	!	!	!	!	!	!	!	!	"-^#()$!	!
K(!012345!"67%$!8!	!	!	!	!	!	!	!	!	!	!	!	%^-(*__!
>?3;62@4!	!	!	!	!	!	!	!	!	!	!	!	"-^)+-$!
G@461@?=!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!	9:=3??43!
TU=31V:62@4=!	%(,!	%(,!	%(,!	%(,!	%(,!	%(,!	%%,!	%%,!	%%,!	%(,!	%(,!	%(,!
H=3P5@!B7=WP:135!	-^%&*!	-^%++!	-^%(%!	-^%&)!	-^%*'!	-^%()!	-^&-,!	-^&''!	-^&%+!	-^%'-!	-^%,*!	-^%()!

X@630!53M345346!V:12:U?3!2452;:63=!:4!2YM1@V35!53U6!1:624C!6<1@PC<!@V311P?24CZ!;@45262@4:?1?@C2=62;!13C13==2@4=!6<:6!;@461@?!0@1!13C2@4!02\35!3003;6=Z!13M@1635!:13!6<3!;@3002;2346!3=62Y:63=!@0!6<3!13C13==2@4=Z!=6:45:15!311@1=ℕ!1@UP=6!6@!;?P=63124C!:6!6<3!13C2@4!?3V3?ℕ!24!M:1346<3=3=^!9:=3??43!;@461@?=!24;?P530!913:;<3=ℕ!GHF I ℕ! A3;<:42;:?!B:624C0!J2C<ℕ!K L H!M31!;:M26:!"?4ℕ!67#$ℕ!K L H!C1@O6<!"67#$ℕ!H@MP?:62@4!"?4ℕ!67#$^! I ??!13C13==2@4=!24;?P53![3:1!02\35!3003;6=^!/2C4202;:4;3!?3V3?=0!_!M`^#-ℕ!__!M`^-(ℕ!___!M`^-#!

!

!

H1@b3;62@4!	! TU=31V:62@4=! "?306!:45!12C<6!@0!;P6@00$!	9:45O26<!V:?P3=! "?306!:45!12C<6!@0!;P6@00$!	M7V:?P3!
Hc!@0!53U6!]!KLH!	! ,#-Z!#&-!	)^)'%Z!+^&&*!	^&)'!
Hc!@0!53U6!]!>\M@16=!	! +*%Z!#)+!	&-^+,'Z!&&^(%,!	^-++_!
Hc!@0!53U6!]!B3V34P3!	! #-%*Z!#&!d!	&-^'#&Z!&(^+)+!	^-()_!
L3U6!/31V2;3!]!>\M@16=!	! ,,&Z!'*!	&^&-%Z!&^*-+!	^'#-!
L3U6!/31V2;3!]!B3V34P3!	! #-&-Z!#-!d!	&^&('Z!&^+,*!	^-#+__!

X@63=0!13C13==2@4!52=;@4624P26[!Y:42MP?:62@4!63=6!P=24C!?@;::?!M@?[4@Y2:?!534=26[!3=62Y:62@4Z!
M7V:?P3!;@113=M@45=!6@!:!63=6!@0!6<3!4P??!<[M@6<3=2=!6<:6!6<3!534=26[!2=!4@6!52=;@4624P@P=Z!
d!63=6!13M@16=!:!O:1424C!6<:6!6<3!U:45O256<!Y:[!U3!6@@!?@O!U3;:P=3!@0!6@@!03O!@U=31V:62@4=!"X12C<6!`!%-$Z!
=2C4202;:4;3!?3V3?=0!_!M!`!^#-N!__!M!`!^-(!

!

!

Appendix

!	"#$!	"%$!	"&$!
A@531:63!B2=E!	7-^-#-_!	7-^-#(__!	7-^-%-__!
!	"-^--($!	"-^--($!	"-^--*$!
J2C<!B2=E!	7-^-##!	7-^-%&__!	7-^-%&__!
!	"-^--*$!	"-^--+$!	"-^-##$!
KLH];:M26:!"?4$!		7-^--*!	7-^--(!
!		"-^--'$!	"-^--'$!
KLH!C1@O6<!		7-^--%__!	7-^--%__!
!		"-^--#$!	"-^--#$!
H@MP?:62@4!"?4$!		7-^--(__!	7-^--(__!
!		"-^--%$!	"-^--%$!
A3;<:42;:?!B:624C0!A@531:63!B2=E!			-^-#%!
!			"-^--*$!
A3;<:42;:?!B:624C0!J2C<!B2=E!			-^--'!
!			"-^-#-$!
Q3:1!R>!!	Q3=!	Q3=!	Q3=!
B3C2@4!R>!!	Q3=!	Q3=!	Q3=!
TU=31V:62@4=!	&'*!	&''!	&'#!
B7=WP:135!	-^#((!	-^%%)!	-^%&+!

X@63=0!TS/!02\35!3003;6=!13C13==2@4=N!53M34534 6!V:12:U?3!2=! e @1?5!9:4E!?34524C!"a KLH$!24! [3:1!6f#Z!1@UP=6!=6:45:15!311@1=!24!M:1346<3=3=Z!=2C4202;:4;3!?3V3?=0!_!M`^#-N!__!M`^-(N!___! M`^-#!

Variable	Obs.	Mean	S.D.	Min	Max	Source and description
913:;<3=!	367!	7.79!	12.83!	0!	85.56!	FAR!L/R!L:6:=36Z!
						M31;346:C3!@0!U13:;<35!53U6!6<13=<@?5=!
GHFI!	364!	3.37!	0.43!	2.43!	4.36!	FAR!L/R!L:6:=36Z!
						G@P461[!H@??;[!:45!F4=626P62@4:?!I==3==Y346!
GP11346!I;;@P46!	370!	-0.09!	0.12!	-0.55!	0.43!	FAR!"%-#):$Z!
9:?:4;3!"aKLHN!67#$!						U:?:4;3!@4!;P11346!:;;@P46!"aKLH$!
L3U6]KXF!"67#$!	353!	41.91!	34.23!	4.29!	457.8!	e@1?5!9:4E!"%-#)$Z!
						3\6314:?!53U6!=6@;E=!"a!@0!KXF$!
L>SR!6@!./!	370!	0.8!	0.15!	0.52!	1!	g@?@!"%-#%$Z!
						52=6:4;3!:5bP=635!36<4@7?24CP2=62;!01:;62@4:??h:62@4!2453\!"L>SR$!
L/R!1:624C!2YM1@V35!	367!	0.23!	0.42!	0!	1!	FAR!L/R!L:6:=36Z!
						2452;:6@1!i!#!20!Y3;<:42;:?!12=E!1:624C!2=!<2C<31!6<:4!@002;2:?!1:624C!
>?3;62@4!	370!	0.23!	0.42!	0!	1!	93;E!36!:?^!"%--#$N!PM5:635!O26<!240@1Y:62@4!@4!3?3;62@4!5:63=!01@Y!
						e2E2M352:Z!2452;:6@1!i!#!20!?3C2=?:62V3!@1!3\3;P62V3!3?3;62@4!6@@E!
						M?:;3!24![3:1!6!
>?3;62@4!")!Y@46<=!	370!	0.15!	0.36!	0!	1!	93;E!36!:?^!"%--#$N!PM5:635!O26<!240@1Y:62@4!@4!3?3;62@4!5:63=!01@Y!
?:631$!						e2E2M352:Z!2452;:6@1!i!#!20!?3C2=?:62V3!@1!3\3;P62V3!3?3;62@4!6@@E!
						M?:;3!24!6<3!=2\!Y@46<=!0@??@O24C!6<3!MPU?2;:62@4!@0!6<3!L/I!
>11@1!24!FAR!K1@O6<!	358!	0.39!	3.93!	-9.17!	34.33!	FAR!"%-#):$!52003134 6!V246:C3=Z!
R@13;:=6=!						K1@O6<!1:63!M1@b3;62@4!0@1![3:1!6!5@43!24![3:1!67#N!Y24P=!:;6P:?!
						KLH!C1@O6<!24![3:1!6!
KLH!C1@O6<!"67#$!	368!	4.84!	4.85!	-37.01!	21.02!	e@1?5!9:4E!"%-#)$Z!
						C1@O6<!@0!KLH!
KLH!M31!;:M26:!"?4N!67#$!	366!	7.02!	0.82!	5.37!	8.99!	e@1?5!9:4E!"%-#)$Z!
						4:6P1:?!?@C:126<Y!@0!KLH!M31!;:M26:!24!;@4=6:46!%--(!./jD!!

Variable!	Obs.!	Mean!	S.D.!	Min!	Max!	Source and description !
JFHG!"LH$!	370!	0.54!	0.5!	0!	1!	FAR!L/R!L:6:=36Z!
						2452;:6@1!i!#!20!;@P461[!M:162;2M:635!24!6<3!<3:V2?[!2453U635!M@@1!
						;@P46123=!24262:62V3!:45!13:;<35!:6!?3:=6!6<3!k53;2=2@4!M@246l!
FAR!M1@C1:Y!!						