



Munich Personal RePEc Archive

Rule Bending in International Organizations: Explaining Instability in the Stability and Growth Pact

Baerg, Nicole Rae and Hallerberg, Mark

University of Mannheim, Hertie School of Governance

31 October 2014

Online at <https://mpra.ub.uni-muenchen.de/18084/>

MPRA Paper No. 18084, posted 10 Nov 2014 07:52 UTC

Rule Bending in International Organizations: Explaining Instability in the Stability and Growth Pact

Nicole Rae Baerg

Mark Hallerberg

University of Mannheim

Hertie School of Governance

nicole.baerg@uni-mannheim.de

hallerberg@hertie-school.org

November 3, 2014

Abstract

In this paper, we consider how European institutions contributed to the euro crisis. In principle, the Stability and Growth Pact was intended to minimize externalities by preventing macro-economic “bad behavior” in the form of large budget deficits. In practice, it has had a difficult history, with several Member States running “excessive deficits” and the Pact clearly failing to prevent the crisis. An important part of the story, and one that is often ignored, is that Member States not only broke the rules but repeatedly bent them by augmenting and amending the European Commission’s assessments. We operationalize what it means to bend rules under the Pact, and we consider explanations for why some Member States repeatedly bent the rules in the run-up to the crisis while others did not. Using a new dataset of Commission assessments of member state economic programmes and Council of Minister revisions of those assessments for the period 1998-2012, we find that big states and states with euroskeptic populations regularly undermined the “watchdog” function of the Commission. The evidence leads us to conclude that unlike some domains where rule flexibility leads to better, and deeper cooperation, such flexibility eroded cooperation in the run-up to the euro crisis.¹

¹The authors would like to thank William Roberts Clark, Mark Copelovitch, Jeffry Frieden, Mareike Kleine, Will Lowe, Robert Thomson, Stefanie Walter, and participants at the 2014 PEIO conference. Research Assistance was generously provided by Andrei Ilas, Kyle Ott and Grzegorz Wolszczak. We also want to especially thank Alyssa Taylor for excellent research assistance.

1 Introduction

This paper examines the operation of the European Union’s Stability and Growth Pact (SGP) both prior to, and during, the euro crisis. While “growth” is part of the title, its primary goal is to keep deficits low to minimize the likelihood that a fiscal problem in one Member State affects another. The SGP has a “preventive” and “corrective” mechanism. The “preventive” arm requires every European Union Member State to submit annual macro-economic programmes to a supra-national body, the European Commission, for evaluation. Those programmes include information on deficits, debts, and the economic growth trajectories of Member States. The Commission, in turn, writes a report on each Member State that constitutes draft text for the intergovernmental Council of Economic and Finance Ministers (ECOFIN, or Council) for a formal statement to the Member States. The key benchmark of the SGP is that Member States should have a budget deficit of 3 percent or less. In this paper, we refer to the Commission’s power to comment on national programmes and make concrete recommendations as its “watchdog” function. The Commission may also recommend corrective measures. The Council, based on an initial recommendation from the Commission, may deem a Member State with deficits above this figure an “excessive deficit” country. If a Member State ignores the Council’s ruling, the Council, again with initial Commission recommendation, can decide to fine a non-complying country. In practice, in the lifetime of the Pact to date (1998 through 2014) the Commission has not recommended fines and the Council has not imposed them. An important reason why there have been no fines, we argue, is that Member States succeed in blocking critical text against them from the Commission early in the process. The “preventive” arm sets up the “corrective” arm; no action against a Member State can begin unless the Council identifies clear issues first.

In terms of actual performance, the Stability and Growth Pact did not succeed in keeping budgets “close to balance or in surplus” since its introduction in 1999. Two large founding members of the European Community, France and Germany, were especially visible in their violation of

the Pact in 2002 and 2003. The Council repeatedly ignored the Commission's advice, and the Commission took the Council to the European Court of Justice. The outcome of the court case did not materially affect the overall institutional setup— it still left the member states themselves, voting through the Council by qualified majority, to decide whether to impose the Commission's recommendations.

Given this background, some have considered the Stability and Growth Pact an institutional failure because it was unable to restrain the fiscal behavior of Member States.² Perhaps even more interesting, however, is whether or not this institutional setup contributed to the onset of the euro crisis. Was the institutional architecture ineffective? If so, did the review process of the Commission's recommendations in the Council contribute to this ineffectiveness? This paper examines these questions.

Recent work in international organizations that examines how institutional design is related to compliance and the enforcement of rules is instructive for these questions. Scholars have shown that there are significant benefits to flexibility in rules, especially in the domain of trade policy. The argument is that, under conditions of uncertainty, countries have a hard time committing to rules. A lack of flexibility constrains countries even further and, as a result, countries only agree to shallow commitments. Alternatively, rule flexibility allows countries to make deeper commitments, which solves their commitment problem.³ In the European context, scholars have also indicated that domestic institutions can both help and hinder compliance with EU level rules.⁴ Moreover, as recent work suggests, institutional design can contribute to economic crises;⁵ this research finds that politics, and especially state power, can exacerbate crises in a multilateral context.

As we show below in our focus on the “preventative” part of the Pact, some Member States

²De Haan, Berger and Jansen (2004) and Konrad and Zschäpitz (2010)

³ Kucik and Reinhardt (2008); Mansfield, Milner and Rosendorff (2002)

⁴ Heipertz and Verdun (2010); Koehler and König (forthcoming); König and Mäder (2013); Richard and Caraway (2014)

⁵Chapman et al. (Working Paper); Dreher and Vaubel (2004)

bent the rules. We document large variation in the Council editing of Commission evaluations of Member States across countries and over time. Unlike for international organizations established to regulate trade, flexibility in the SGP did not provide Member States assurances of deeper commitment to the framework. Instead, it made it easier for some states to weaken the operation of the Pact. In this paper, we argue that by bending the rules, Member States endogenously eroded the power of the European Commission in its role as a “watchdog.” Thus, unlike the trade literature that finds that rule bending is beneficial for contracting, in the case of the SGP rule bending weakened the Member State commitments to the EU and made it easier for powerful Member States to pursue uncooperative strategies. Furthermore, we argue that these uncooperative strategies contributed to the crisis. Finally, similar to studies that emphasize power,⁶ we find that one important predictor of rule bending is Member State power, which in this case is voting power on the Council of Ministers.

To examine whether the institutional configuration of the EU, and in particular, the de facto flexibility of the SGP contributed to the euro crisis, this paper utilizes a new dataset constructed from human coding and computer textual analysis of existing Commission reports on Member State Stability and Convergence Programmes and the changes the Council of Ministers made to them. Using this new data, we examine whether economic performance, Member State power, audience costs, domestic institutions, and log-rolling help explain whether the Council changed the Commission’s recommendations. As a preview of our results, we find that countries with the largest voting power on the Council (10 votes prior to the Nice Treaty, 29 votes after) are more likely to receive favorable edits to the Commission’s original text. Also interesting is that countries with relatively euroskeptic domestic populations are also more likely to have the Council change the Commission’s original judgment against it. We also find that countries are more likely to bend the rules when rule bending is prevalent. We conclude from our analysis that powerful Member States and Member States with euroskeptic domestic populations were able to receive the necessary

⁶Chapman et al. (Working Paper); Copelovitch (2010)

number of votes in the Council to undermine the watchdog function of the European Commission, which undermined institutional oversight mechanisms that may have lowered the debt levels of several member states especially during “good times” and left them less exposed once the crisis was underway.

The paper proceeds as follows. The first section examines the dynamics between the Commission and the Council. It explores the Commission’s decisions on member states and the Council’s decision whether to weaken the Commission’s recommendations. The second section explores the intensity of editing that Member States made to the Commission’s evaluations of Member State economic programs. While the dataset comes from the European Union, the analysis has broader theoretical relevance. Work in international relations considers why and when flexibility can deepen commitment and when it cannot. Similarly, work in international relations considers how Member State power contributes to economic crises. We find that in the context of the SGP, Member States behaved in a way more akin to traditional models of power politics and audience costs than models of deepening commitment and cooperation. We argue that the behavior of the Member States, especially large and euroskeptic Member States, weakened the oversight mechanism of the Commission over time.

2 The Stability and Growth Pact

Before explaining Member State behavior under the Stability and Growth Pact, it is necessary to explain how it worked and which actors were involved. The Maastricht Treaty provided the road map to create a common currency and an independent central bank to regulate it. While fiscal policies were to remain the domain of Member States, the Treaty anticipated the need for some sort of macro-economic policy coordination. As Article 99(1) of the Treaty stated, “Member states shall regard their economic policies as a matter of common concern and coordinate them in the Council.”

The European Monetary System crisis in 1992-1993, combined with a general worsening of budget deficits across the Union, heightened concerns about how Member States would perform once the Economic and Monetary Union began. Two of the so-called Maastricht criteria focused on fiscal policy and set reference values of a general government deficit no larger than 3% of GDP and general government debt burden no larger than 60% of GDP. The Commission used these values, in turn, to judge whether countries met the requirements to join the Euro Area (Stage III). In practice, the emphasis in the run-up to the introduction of the euro was on the deficit figure, and Member States had to get their deficits at or below 3% if they wanted to adopt the euro.⁷ However, there were only vague provisions for what to do once the euro was created.

Article 103 of the Maastricht Treaty was judged as a “no bailout clause” because it stated that neither the Union level nor the other Member States were liable for the commitments of other governments. There was concern, however, that this was not a credible position, and indeed, the financial packages for Greece, Ireland, and Portugal in 2010 and 2011 confirmed this fear. A state with a fiscal crisis could hurt all members of the Euro Area, and the remaining countries could subsequently find it in their best interest to assist this state. Consequently, there seemed little to prevent states from relaxing their fiscal stances once they joined.

These concerns led to the creation of the Stability and Growth Pact, which the Member States agreed at the Dublin Summit in December 1996.⁹ The details of the Pact’s operation are important in assessing whether the Pact has been successful in improving fiscal discipline and in coordinating the fiscal policies of the Member States. At its core, it has both preventive and corrective mechanisms. The preventive mechanism focuses on detailed monitoring of what Member States

⁷The evaluation of the relative “fitness” of Member States for Stage III of EMU gave the Commission some agenda-setting power. The Treaty indicated that a “ratio that has declined substantially and continuously and reached a level that comes close to the reference value” (Article 104(2a)) would suffice. In 1994, the Commission judged that all countries but Luxembourg and Ireland had “excessive deficits.” While Ireland had a debt level well above 60%, the Commission argued that the ratio was going in the proper direction. This meant in practice that countries such as Belgium and Italy with debt levels above 100% of GDP could qualify for Stage III of EMU so long as their deficit number was below 3%. ⁸

⁹See [Heipertz and Verdun \(2010\)](#) for more detail about the Pact’s creation.

are doing. While states submitted some information to the Commission beginning in 1994, the contents of those reports were not well-specified. States would have to write either convergence (non-euro members) or stability programs (euro members) and update those programs roughly at the end of each year. All states were expected to have budgets that were “close to balance” or in surplus over the medium-term. This means that they were to adjust their balances according to the economic cycle—in good economic times they were to run surpluses, while in bad times they could run deficits. The Commission then had the responsibility to review each programme, and to make recommendations to the ECOFIN Council on whether the programme met European goals and whether the goals set in the programmes were realistic, given domestic conditions. The Commission could also recommend that a state receive an “early warning” that it would run an excessive deficit if immediate action were not taken, which the ECOFIN Council would subsequently have to approve in order for the warning to become official. ECOFIN would have to approve the Commission’s recommendation by qualified majority vote of all Member States, which meant that “countries that ‘sinned’ retained the right to vote and needed only a few additional countries prospective sinners among them to block such steps.”¹⁰

The corrective arm followed from the Commission’s assessment of a given state’s domestic policy. If it judged both that an excessive deficit existed and that a state had not made any progress in eliminating it, the Commission could recommend to the Council that the state make a non-interest-bearing deposit with it of up to 0.5 percent of GDP, depending on the size of the deficit. If the Member State continued to neglect necessary reforms, the deposit could become a fine.

Perhaps the biggest shock to the Pact in the initial years was the behavior of two of the founding members of the European Union, France and Germany. They had problematic deficits beginning in 2002. The Commission recommended an “early warning” to Germany in 2002, but the ECOFIN Council did not pass it, and an informal agreement reached with Germany meant that Chancellor Schröder avoided getting such a warning in a national election year. In autumn 2003, the Commis-

¹⁰[Schuknecht et al. \(2011, pg. 9\)](#)

sion judged that the deficits in France and Germany were excessive and that they would continue, and it attempted to begin the sanctioning process against each Member State. The ECOFIN Council, however, balked at following the Commission's recommendation. The Commission's response in January 2004 was to take ECOFIN to the European Court of Justice. In its judgment, the Court ruled that the Council could not suspend a procedure without the Commission's recommendation,¹¹ but this was a hollow victory—no additional action was taken against France and Germany.

There was a clear sense both in the academic and policy communities that something was wrong with the Pact, resulting in a slew of reform proposals.¹²¹³ The Member States themselves agreed to a concrete reform in March 2005. Under the revised Excessive Deficit Procedure (EDP), more factors were explicitly considered when determining whether a country had an excessive deficit, such as if spending fostered international solidarity or if it promoted the unification of Europe.¹⁴ Member States also set their own medium-term objectives, which could include the future fiscal effects of major structural reforms in their calculations. Once a state was found to have an excessive deficit, it could receive extended deadlines so that it had a longer period to correct the deficit than before the revision. Countries experiencing negative growth were exempt, whereas the previous requirement for automatic exemption was a decline in economic output of 2 percent of GDP.¹⁵ In summary, the Pact was weakened.

A further reform came during the euro crisis. The so-called “Six Pack” routinized the “preventive” part of the SGP through a new “European Semester.” It also put more emphasis on a debt level of 60 percent or below, not just on the 3 percent deficit target. Perhaps most impor-

¹¹ [Morris, Ongena and Schuknecht \(2006, pg. 17-8\)](#)

¹² see [Begg et al. \(2004\)](#); [Schelkle \(2005\)](#); [Schuknecht \(2004\)](#)

¹³ The best summary of the proposals appears in [Fischer, Jonung and Larch \(2006\)](#) appropriately-titled paper “101 Proposals to Reform the Stability and Growth Pact. Why so Many?”

¹⁴ The clause on “international solidarity” was widely seen as a way to mollify France, which wanted defense spending to be exempt, while “European unification” would allow Germany to include the costs of German unification.

¹⁵ An excellent review of the changes in the Pact appears in [Morris, Ongena and Schuknecht \(2006\)](#)

tantly, it changed the decision-making rules—rather than a need for a QMV in support of what the Commission proposes, it would then take a reverse qualified majority vote (RQMV) to block a Commission proposal or recommendation. These changes entered into force only on 13 December 2011, however, and we do not have sufficient data to judge whether these changes had an effect on the decisions we document here.

In summary, the process under the Stability and Growth Pact is more flexible than a simple 3 percent deficit rule, and yet, the core procedure has remained the same through 2014. The Commission evaluates every Stability and Convergence Programme and issues a report where it can point out where it finds that the given Member State has an issue. These texts, however, are not the final decision—the ECOFIN Council has to approve each one by qualified majority (through 2011), and the Council could (and did) change the Commission’s text. Conveniently for our purposes, the various texts at the different stages of the process are publicly available. In fact, the debate between the Council and the Commission are embedded into textual changes and edits, which makes systematic analysis of the texts useful for extracting data.

3 Explaining Rule Bending

The European Commission evaluates the macro-economic policies and macro-economic performance of every government. Once the evaluation is complete, the Commission reports the evaluation in the form of a textual document and gives it to the Council. For this paper, if the Council accepts the Commission’s evaluation unchanged, then we say that the Council (on behalf of the Member States) is abiding by the rules. Rule bending, by contrast, is when the Council (as the representative of the Member States being evaluated) changes the textual evaluation of the Commission. When the Council changes the text of the evaluation, we say that there is rule bending; this is because the Council, by augmenting the Commission’s evaluation, is intervening in the oversight mechanism that the Commission is tasked to perform.

This then begs the question— why would Member States bend the rules? For one, the press as well as the political opposition can highlight and exploit the Commission’s evaluation of the Member States and therefore, Member States may have an interest in making sure that they approve the content. But, as we show in this paper, governments vary in their ability to change the Commission’s textual document and also vary in their sensitivity to what the “watchdog” writes. This paper offers the first look at the determinants of rule bending, as conceptualized by changes to textual evaluations of Member States’ Stability and Convergence Programmes.

We focus on the “preventive” arm of the Pact and we consider “rule bending” as the variation between the Commission assessments and final European Council opinions of the Stability and Convergence Programmes. The empirical section goes into further detail, but we measure rule bending along two dimensions: whether the Council weakened the substantive message of the Commission’s assessment; and the total number of edits required to move from the original Commission text to the one the Council issued to the Member State.

The first explanation for why some Member States receive more favorable changes in the Council relates to economic performance. As explained above, the SGP is meant to take country’s position in the economic cycle, and a country that is performing badly should receive lighter treatment (that is, its economic weakness is the reason for a budget deficit). In addition to the economic explanation, a second explanation is Member State political power. Theories of political power suggest that big states or politically important states tend to ignore treaties, while small and politically less important states are forced to play by the rules. The third argument is that audience costs, either domestic and or international, constrain states even when they are big and powerful. A fourth argument focuses on domestic institutions. A final argument is that there are inter-organizational politics, such as log-rolling, that allow states to behave badly when other more powerful states are behaving badly, but have to rein in their bad behavior when powerful states say so. In each case, we discuss below the logic of the argument, highlighting the empirical predictions for how these factors relate to amending the Commission’s evaluation. The following section then offers

an empirical test of these theories.

3.0.1 Economic Performance

One simple economic explanation for why a Member State would change the Commission's recommendations has to do with economic performance. From a purely technocratic perspective, if the Council was concerned only with the Member State's economy, then we would expect to see a positive relationship between economic success and editing.¹⁶ A simple economic explanation suggests that Member States are more likely to amend when Member State economies are doing well and the Commission, which is charged with enforcing what amounts to a structural budget rule, is complaining that governments should save more of the proceeds from growth. Therefore, empirically this implies that we should find a positive correlation between a economic performance and rule bending.

3.0.2 Relative Power

Political considerations probably add to strictly economic ones. One way to view the behavior of states under the Stability and Growth Pact is through the theoretical lens of collective action.¹⁷ One might conceive of EMU as a group of states cooperating to provide the collective good of a common currency. Olson argues that in unequal groups above a certain size, it typically falls to the largest member(s) to bear the burden of providing the collective good. While one could argue that Germany's actions to preserve the euro during the crisis fit this type of explanation, it does not explain why Germany is one of the states to shirk under the institutional device meant to maintain the public good, namely the SGP.

In understanding this puzzle, the most simple political explanation may be relative power.

¹⁶An analogous discussion of the economic view in IMF lending is presented in [Copelovitch \(2010\)](#). If only economic concerns matter then we should not see political variables exerting an influence on member state editing at all.

¹⁷[Olson \(1965\)](#)

When examining multilateral lending and conditionality, scholars have found that powerful countries exert significant influence¹⁸ and that state power shapes the size and the conditionality of loan disbursements.¹⁹ As Stone observed in his discussion of France and Germany's behavior under the Stability and Growth Pact, "the rules of the Pact are difficult to enforce against powerful countries."²⁰ Applying this logic to the SGP, powerful states might be more likely to bend the rules because the potential punishment they face is less than punishments faced by small states. That is, what can the European Commission do if Germany wants to amend the Commission's recommendations? In contrast, if Portugal receives specific recommendations for adjustments to its budget policy, it may be more vulnerable to pressure from other states.

There is a more direct way that country size matters in our analysis, namely through the voting rules, which generally provided large states with outsized actual power to change the text. There are four separate voting regimes over our time period. In the first, which runs 1998-2002, it takes a qualified majority vote (QMV) to pass the Commission's text on a given Member State, with qualified majority representing with QMV representing 62 of 87 votes and Member States having between 2 and 10 votes. The four largest states by population each had ten votes. The Treaty of Nice introduced a so-called "triple majority" as the definition of QMV. As Tsebelis and Yatahanas notes, however, this had the effect of strengthening the Council vis-a-vis the Commission because it increased the threshold needed for the Council to get to a qualified majority. It became somewhat easier for a large state to block a given Commission proposal.^{21 22}

¹⁸Thacker (1999)

¹⁹Copelovitch (2010)

²⁰Stone (2011, pg. 118)

²¹Tsebelis and Yatahanas (2002)

²²One difficulty we face is that formal votes where there are divisions on such documents are extremely rare, and the discussions of them in ECOFIN are secret. At the same time, we have been told off the record that Member State representatives routinely prepare changed text for these meetings. We expect that we are picking up the sensitivity of governments to the Commission's comments, and whether the Council then passes this text, in the observed differences in the documents, we measure and report here.

One should note that there could be strategic interaction between the European Commission and the Member States through their representation on the Council of Ministers. A Commission that anticipates big state dominance may be reluctant to recommend specific actions for the large states in the first place. In this case, the Commission would make recommendations only for small states. Empirically, however, we also see large states have the Commission's texts revised and weakened. While this begs the question of whether large states would get even stricter recommendations if they were not so powerful, it suggests that we would get a null or negative sign on the coefficient in our analysis—large states do not need to ask for changed Commission texts for them because the Commission would have written texts the large states wanted from the beginning.^{23 24}

3.0.3 Population Preferences

A literature in international relations explains how general populations can constrain governments. Audience costs are those governments suffer when they do not live up to commitments they have made. The audience can be international or national. At the international level, Finlayson and Zacher contend that the General Agreement on Tariffs and Trade (GATT) provided a forum for states to observe each other's behavior. Governments that choose to cheat under the Treaty acquire a reputation for not honoring agreements, and other governments would be reluctant to cooperate with them in the future.²⁵

This argument has been extended to the domestic level. Mansfield, Milner and Rosendorff find that democracies are more likely to sign and, later, to comply with international trade agreements

²³There is evidence that the Commission engages in such strategic behavior in cases it brings to the European Court of Justice. That is, it presents the cases it expects to win. See [Carrubba \(2005\)](#)

²⁴This is the baseline model, and if it explains everything other possible explanations are not relevant. This goes in particular for the argument that interaction in the Eurogroup among ministers constitutes peer pressure, and if there is enough pressure on offenders in this off-the-record arena then the offenders will reform their ways. We would like to emphasize that we know of no academic piece that makes this argument, but discussion in Brussels and elsewhere on the utility of the Eurogroup focuses on peer pressure to resolve disputes. It may very well be, of course, that peer pressure is useful for smaller issues, but that the more visible compliance or non-compliance under the Stability and Growth Pact would not be subject to peer pressure.

²⁵[Finlayson and Zacher \(1981\)](#)

because of domestic audience costs. In this model, government leaders who violate the terms of an agreement suffer a loss of reputation in the minds of voters; these voters, in turn, are inclined to vote out the offending government at the next election.²⁶ Similarly, Jensen suggests that democracies create audience costs for leaders who might wish to expropriate property of foreign firms.²⁷ He argues that expropriation leads to a reputation loss in the minds of international market actors, who would then seek to undermine the ruling government (withdraw investment, etc.). Voters, for their part, may then observe drops in foreign direct investment and vote out the incumbent.

Despite being used in a variety of subject areas with variation in the type of audience, all of these models share general traits. First, they assume that an informational signal is clear to the audience. Second, they make an implicit assumption that the audience cares, i.e., that it prefers compliance over defection. Lastly, they assume that the audience can, and will, act in such a way that may deter the government from defection in the first place.

Using experimental research methods, Tomz provides micro-level evidence that individuals do, indeed, care about abiding by the terms of international agreements.²⁸ He demonstrates that citizen preferences for policies are conditioned by whether or not particular policies are sanctioned by international law/agreements. While much of the literature on audience costs speaks to the reputational costs of reneging (and potential impact on individuals' material interests), Tomz allows that the mechanism of audience costs may simply be individual norms, i.e., that populations think it is wrong to cheat on agreements.

Arguments about audience costs usually assume that citizens want their governments to honor their treaties. Chaudoin, however, finds in a study using a survey experiment that voters worry about keeping to a treaty only when they do not have strong preferences over a given policy: a

²⁶Mansfield, Milner and Rosendorff (2002)

²⁷Jensen (Working Paper)

²⁸Tomz (2006)

voter who opposes free trade would prefer that her government not comply under a trade treaty.²⁹

In the European Union’s case, we know that voters have different opinions across Member States on the Treaty itself. Some, such as Ireland, have populations that were, on average, euro-optimists (at least before the crisis), while others like the United Kingdom had a more euroskeptic population. Because the treaty here relates to the project of European integration, we expect that Member States with citizens who are generally euro-optimists are more likely to want leaders to follow what the European Union says. In fights with the European Commission, a population may be more likely to side with the Commission where there is enthusiasm for European integration. In contrast, in euroskeptic Member States, we expect the reverse—governments are more likely to bend the rule.

3.0.4 Government Preferences

Under what circumstances would a Member State government not want to accept the Commission’s ruling?

Governments are more sensitive to what populations think of their performance and may not want Commission criticisms of them prior to elections. Moreover, there may be more “bad” behavior for the Commission to complain about in an election year, as governments may try to boost the domestic economy by spending more and/or cutting taxes. Heipertz and Verdun, in particular, argue that electoral considerations were especially prominent in the original design of the Stability and Growth Pact and so they may also matter in explaining textual changes to the Commission’s recommendations.³⁰ This is the domestic story we explore in our main set of results, with the expectation that there is more editing of Commission texts in election years.

There are additional domestic factors we consider in the robustness section. There is a developed literature on Member State compliance under European law. The predictor for behavior

²⁹Chaudoin (2014)

³⁰Heipertz and Verdun (2010)

in many of these studies is the position the Member State took prior to passage of the given law. Thomson, Torenvlied and Arregio consider the impact of the negotiation stage on the implementation stage, with the focus on directives and subsequent infringements and delays.³¹ Such directives must be transposed into national legislation. They find no evidence of country-specific effects, but they do find that countries that opposed the directive during its passage are more likely to face an infringement case. Similarly, König and Luetgert, in a large n quantitative study, find that countries that fail to transpose their directives on time were more likely to have been opposed during passage.³²

Given that France and Germany pushed through the Stability and Growth Pact, the previous literature would suggest that these Member States would be more likely to comply. Yet these countries were widely seen as the first to break the Pact. This behavior suggests that country positions in the mid-1990s when the Stability and Growth Pact was originally passed are not a domestic reason why a government would be more likely to push for changes in Commission assessments. At the same time, the prediction would be the opposite of the one based on the size of the state, that is, France and Germany should be more likely to comply, and as a robustness test later in the paper with explore this argument in more detail.

Another explanation of domestic government behavior is partisanship. Strict budget balances could be seen as something that the political right weights more than the political left. Empirical work on actual performance of Member States would dispute this however– it finds no partisan effects on the size of deficits or changes in gross debt burdens.³³ There may be partisan differences across governments concerning the desirability of meeting the targets in the Pact that are worth exploring, especially relating to the preferences of the finance minister as well as the prime minister, and we explore this argument further in the “Robustness” section.

³¹Thomson, Torenvlied and Arregio (2007)

³²König and Luetgert (2009)

³³Hallerberg, Strauch and von Hagen (2009)

3.0.5 Log-rolling

Another explanation might be that Member States are more likely to bend the rules when other Member States are doing the same, or log-rolling. Again in the case of multilateral institutions, scholars indicate that lax enforcement of rules can lead to bad behavior whereas stricter enforcement of rules can lead to good behavior.³⁴ The average number of infringements, measured as weakening and editing, as well as the euro crisis gives us some leverage for examining this hypothesis. One way that we can test this is by looking at the average amount of substantive weakenings over time. In the empirical section we include a measure of the average number of changes to the evaluations, either weakening or editing, to capture whether or not Member States are more likely to weaken when others are weakening and editing more intensively. Furthermore, we might also expect that log-rolling might be different before and after the Euro Crisis. During the Euro Crisis, Member States were encouraged by the European Commission to run larger deficits throughout the crisis than before the crisis. In fact, deficit spending was encouraged by the Commission and, in 2009, the Commission suggested that Member States should generally spend money to combat the crisis. Therefore, we expect to see a positive relationship between the average number of weakening and edit intensity with weakening and intensity, and we also expect to see less oversight during the crisis and therefore less editing and weakening at the end of the sample period.

In summary, Member States submit either convergence or stability programmes yearly to the European Commission. The Commission, in turn, evaluates the programme and submits an assessment to the ECOFIN Council. The Council then decides whether to accept the Commission's evaluation as-is, strengthen the evaluation, or weaken it. In addition to the direction of changes, the Council also decides with what intensity to edit the Commission's text. Finally, once handed the new text by the ECOFIN Council, the government then decides whether to comply with an explicit recommendations. Given the theoretical expectations outlined above, and given that we are inter-

³⁴Copelovitch (2010); Dreher and Vaubel (2004)

ested in the determinants of rule bending in the SGP even before Member States decide whether or not to comply with the Commission's recommendations, in the next section, we evaluate the following expected relationships:

Economic Performance: Member States with strong economies will receive more edited, and weakened, text than member states that are performing less well.

Political Power: Large member states will receive more edited, and weakened, text than smaller member states.

Audience Costs: Countries with populations that are more euroskeptic will receive more edited, and weakened, text.

Domestic Institutions: Member states in an election year will receive more edited, and weakened, text.

Log-rolling: Member States will receive more edited, and weakened, text when others receive such weakened, and edited, texts.

4 Measurement and Analysis

4.1 Dependent Variable

We examine variation between the European Commission assessments and European Council opinions of the Stability and Convergence Programmes between 1998-1999 and 2012-2013. The countries in the study consist of the EU-15.³⁵ One might wonder why we include the three countries that chose not to join EMU, namely Denmark, Sweden, and the United Kingdom. We see these states as potentially interesting sources of variation. They must submit convergence programmes

³⁵Countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and United Kingdom

yearly, and those programmes undergo the same review as the stability programmes that the Eurozone members face. As “outs,” they are not subject to the same possible sanctions that Eurozone members face, and in particular, they are not subject to possible fines, so it may be that they do not face the same incentives. By including them in the statistical analyses, we can empirically assess whether or not the strategic behavior of the “in” countries is different from the “out” countries.

While our focus is on the EU-15 because these are the original member states at the beginning of the Stage III of Economic and Monetary Union, which introduced this monitoring procedure for all member states, to be clear that our findings are common after other countries joined, in the robustness section that follows, we report results from an expanded EU-27 sample.

We develop two new measures that account for disagreements between the Commission’s assessments and the wishes and opinions of the Member States, as reflected by the Council. The first measure examines the direction of Member State amendments and reports, specifically whether editing by the Member State made the Commission’s amendments weaker or left it the same. The second measure tracks the frequency that Member States change the Commission’s text. Together, these new measures capture both the direction and also the intensity of amendments.

Using human coding of Member States’ edited Commission documents, we measure the direction of the amendments between the Commission’s assessments and Council’s opinions and put them into two categories: 1. “No change” – there is no noteworthy or substantial variation (defined more precisely below)³⁶ 2. “Weakening” – the Council Opinion has placed less stringent demands or in general is less critical of a country programme than the Commission assessment.

Before one accepts the validity of the measure, a few words about coding are in order. We do not consider statements that are substantively similar yet vary in the level of detail.³⁷ Omissions

³⁶We also coded cases where there was a “strengthening.” These cases were comparatively rare. Moreover, our theoretical focus is on explanations of rule-bending. Another benefit of conceptualizing rule bending on a latent scale is that the direction of rule bending can be assumed to be in the interest of the Member State.

³⁷For example, for Greece 1998-99, the Commission indicated that “structural reform is needed in order to improve the efficiency of the Greek economy, particularly in its large public sector.” The Council’s response was as follows: “The Council welcomes the structural reforms included in the programme which are geared towards the labour market,

in the statement are considered substantively significant if the Commission holds a country to a specific outcome but the Council does not. Treatment of omissions is especially relevant because there is variation in the length and detail in various reports. Finally, we do not record a distinction in the magnitude of weakening or strengthening; if there is a perceived change in one direction or the other in the statement, we simply record it as such. Commission and Council documents are classified as having a weakening, a strengthening, both a weakening and a strengthening, or no change in a given year. Table 1 provides examples from each class.

[Insert Table 1 About Here]

Given these provisos, our first, hand coded dataset includes 224 annual Commission and Council reports on member state programmes.³⁸ Figure 2 “Weakenings by Country” shows the distribution of weakening by Member States over the sample period. There are 67 instances of the Council opinion weakening the Commission assessment. This amounts to approximately 30 percent of the observations, suggestive that rule bending is common.³⁹ Especially striking is the clustering of weakened Council reports. While the Commission’s fight with France and Germany certainly got the most press attention in autumn 2003, most states received lighter Council comments that year than what the Commission first proposed in their original report. Also noteworthy is that at least two states received a weaker report in every year but 2006. The lack of cases of weakening in the most recent years is also noteworthy, especially after the 2011 reforms and the crisis period.

There is also notable variation by Member State. The total number of weakening varies from

the social security system and the wider public sector. The Council urges the Greek Government to implement them as scheduled and pursue further the reform effort in order to enhance the potential and efficiency of the Greek economy.” This case was coded as no significant change.

³⁸Our total dataset has 225 rows with one missing value of weakening for Denmark 2002 and 36 missing values for our second dependent variable, edit distance. We check to make sure that missing data is not skewing our estimates by using multiple imputation in Amelia.

³⁹We also considered whether the Council on balance strengthened, rather than weakened, the Commission text. There are such 28 instances, or about 12 percent of observations. A majority of states have one such case. For the purposes of the analysis that follows, where our theoretical focus is on rule-bending, we treat strengthening cases as “no change.”

a high of 11 (France) to a low of 1 (Denmark, Luxembourg, Ireland and Spain). Collectively, France and Germany had nearly 26 percent of all cases of weakening, despite constituting around 14 percent of the dataset. While this relationship is perhaps unsurprising, the surprise is that many states received lighter comments from the Council. The “out” states of Denmark, Sweden, and the United Kingdom have proportional numbers; they constitute 22 percent of the weakening and they also represent 20 percent of the cases.

[Insert Figure 2 About Here]

While the direction of amendments depicts how Member States oppose the Commission’s evaluations, another equally important measure is the frequency or intensity of editing to the Commission’s evaluations. In order to measure the frequency that a Member State opposes the Commission’s text, our second dataset constructs a metric of the Member State’s editing.⁴⁰ To do this, we use automated textual analysis. First, we find all the instances of textual differences between the Commission and the Council’s text using textual comparison software. Second, examining only the changed text, we calculate the Levenshtein distance metric for each edit made. The Levenshtein distance metric is a measure that computes the minimum number of single word deletions, insertions, or substitutions that are required to match the new text to the old text.⁴¹ For example, if the Council edits the initial recommendation from “2010” to “2010 in view of the uncertainty connected to the economic downturn,” then the Levenshtein distance metric is 10 because there are 10 word insertions necessary to transform the old text to the new text. The Levenshtein distance is commonly used in linguistics, including spelling algorithms, and in statistical models to evaluate language similarity. The textual distance measure is particularly useful here because it picks up word changes in the text, even if small, between the Commission and the Council reports.⁴²

⁴⁰The sample includes the same time period as before with all EU-15 Countries from 1998-1999 to 2012-2013.

⁴¹[Sørensen \(2007\)](#)

⁴²Another possibility is to calculate micro changes, such as changes to characters, or alternatively, macro changes

Second, after calculating the distance metric for each edit per country document (there are for example over 5000 edits in any given year), we sum the total number of Levenshtein distance measures per country-year. Figure 2 also shows the variation in document editing as measured by the Levenshtein distance measure across the sample between 1998-1999 to 2012-2013. In addition to the raw numbers, reported by the dash lines, Figure 3 also illustrates the trend in textual editing and the trend in weakening together on the same plot reported by the solid lines so that we can observe the similarity and the differences between the measures.

One potential problem with the Levenshtein distance measure is that the distance measure does not measure lexical similarity between sentences. For instance, consider two sentences with the same number of words but slightly different meanings such as: “we should expect a 2 percent increase” and “it’s likely that a 2 percent increase.” This example has a calculated distance measure of zero because it has the same number of words and no substitutions have been made. Despite this, however, additional statistical analysis and the correlation of the Levenshtein distance measure with weakening suggests that both measures capture the phenomenon of Member State rule bending that we examine throughout the paper. We also think that the distance measure contributes something important to the study of international organizations as much of what international institutions do is provide a forum for inter-state bargaining over time and this bargaining often produces textual information, which, as we show here, is useful for comparative research.

Like the measure for weakening, editing of Commission documents varies substantially not only across countries but also over time. Countries such Germany, Italy, and France tend to have, on average, higher edit intensity than other, less powerful countries. This pattern is analogous to the observed pattern for weakening. One important source of variation between the measures, however, is the differences in the timing of their peaks; the weakening measure peaks earlier than the edit intensity measure. The benefit of looking at the direction and the intensity of the rule

such as changes to the number of sentences in paragraphs. Measuring word changes is most effective because words can account for small and potentially meaningful additions to text, while simultaneously, not being too sensitive to cosmetic edits such as changes from British to American spelling and punctuation.

bending over time and across countries is that it allows us to capture changes in the use of this particular form of rule bending, and highlights the changing relationship between the Member States as represented by the Council, and the Commission.

[Insert Figure 3 About Here]

4.2 Explanatory Variables

Economic Performance: In order to test whether or not countries' economic performance matters for weakening and editing, we collect data on Member State output gaps. The output gap measures whether countries are performing better (in terms of changes in GDP) than potential output (also measured as change in GDP). Countries that report a positive output gap are growing above potential whereas countries with a negative output gap are growing below potential. Note that this is different than actual growth. For example, Ireland and Germany can have the same growth rate but different output gaps. If Ireland is growing slower than potential, Ireland will have a negative output gap, whereas if Germany is growing above potential, Germany will report a positive output gap even if the countries exhibit the same economic growth rate. The potential growth rates are then subtracted from the actual growth rates. We use this measure of economic performance because the output gap is a key concept in the SGP criteria applied by the European Commission—the SGP is a form of a cyclical budget rule, which means that a given Member State may have a larger budget if its growth is below potential.

Relative Power: We examine whether or not a country is large in terms of vote share they have on the Council. If a country had 10 votes on the Council prior to the Nice Treaty, which is equivalent to having 29 votes after the Nice Treaty, we code this country as “large.” If a country has fewer than 10 votes on the Council, the country is coded as “small.” Furthermore, as discussed previously, we also account for whether or not the country is an “in group” or “out group” member, with Denmark, Sweden, and the United Kingdom coded as “outs” and the other countries coded

as “ins.”

Audience Costs: As a measure of audience costs that relate to the EU context, we measure a population’s expected favorability towards the European Union as a proxy for audience costs. The data that we use is from Euro-barometer. We code the values for each country in response to the question:

Taking everything into consideration, would you say that (your country) has on balance benefited or not from being a member of the European Community (Common Market)?

The question is an appropriate measure of audience costs because it captures the perceived benefits and costs of EU membership.⁴³

Elections and Partisanship: We examine two variables concerning the government’s perspective on the Stability and Growth Pact. To measure the impact of electoral considerations on a government’s fiscal policies, we code for the percentage of a given year that precedes an election.⁴⁴ That is, an election held on July 1 would be coded as 0.5 in the same year and 0.5 in the previous year. As a different specification, we also use the forward version of the variable, which is both an additional control for endogeneity and well as a test for whether member state governments anticipate the effects of a Commission document on later elections.

Another potentially important variable that might help explain the impact of domestic politics on council behavior is partisanship. Grieco, Gelpi and Warren find that shifts leftward make it less likely that governments comply with legal commitments to capital market integration.⁴⁵ Similarly, the budget balance requirements in the Stability and Growth Pact could be popular among right-

⁴³Unfortunately, Eurostat chose not to ask this question in 2012. More on the implications of this for our empirical model appears below. Also note that audience costs in this way assume that costs are born by forward looking Member States who are less concerned with reputation and more concerned with winning a larger share the next election. This seems most appropriate for coalition politics in the EU context.

⁴⁴Franzese (2002)

⁴⁵Grieco, Gelpi and Warren (2009)

wing governments but not among left-wing governments. In order to measure this, we employ Benoit and Laver's left-right partisanship of the Prime Minister and the Finance Minister for all 15 Member States, which appears in (names withheld).⁴⁶

There are a few theoretical concerns with partisanship, however. First, given the presence of coalition governments in most EU-15 countries that often include parties from both the left and the right (for example Belgium and Finland), it is hard to envision a clear qualified majority from one partisan block in place in the Council. Second, the argument does not seem consistent with the evidence— Germany and the United Kingdom were both left-leaning when the European Commission suggested concrete recommendations while France and Italy were both right-leaning. Instead of the partisanship of the government as a whole, however, the partisanship of the Finance Minister may matter. Because the Finance Ministers are responsible for economic policy (and are held accountable for their economic policymaking), it is perhaps more likely that partisan concerns are stronger for the Finance Ministry than the head of the coalition government. In the empirical analysis, we test this directly by using partisanship of both the Prime Minister and the Finance Minister. To offer a preview of the results, we find that neither the partisanship of the Prime Minister or the Finance Minister matters; because of this and our theoretical concerns about the relevance of this variable, we exclude it from the final specifications presented in Tables 1 and 2 (but see the results in the robustness analysis reported in Table 4).

Log-rolling: Finally, in order to measure log-rolling, we construct a spline smoother of the yearly trend for both measures of the dependent variable. A positive coefficient estimate suggests a positive correlation between the likelihood of weakening and editing when other countries are weakening and editing whereas a negative coefficient estimate suggests the reverse.

⁴⁶Benoit and Laver (2006)

4.3 Results

Our first statistical analysis examines the probability of bending the rules as measured by weakening the Commission's recommendations. We want to test whether economic performance, Member State power, audience costs, domestic political institutions, and log-rolling matter for whether Member States weaken the Commission's recommendations.

Because the first dependent variable, weakening, is coded 0 or 1, we use a logit model. As suggested before, we think of the dependent variable as a latent variable that estimates the probability of rule bending, and specifically, bending the rules to make the recommendations less stringent for Member States. The estimates and the standard errors from the logit model are reported in Table 1. The first column reports the findings using a model with year fixed effects and the second column reports the findings using a random effects model with random effects for both year and country. When we estimate the effects of Member State characteristics on the likelihood of weakening, the results show that large Member States and Member States that are euroskeptic are more likely to have the Commission texts written about them weakened. In particular, holding the other variables at their observed values, the probability of a weakening increases from .27 to .41 if one moves from a small to a large state. Thus we find some evidence that politically powerful Member States (as measured by votes on the Council) have more Commission weakenings.

Another striking result is that the more that the Member State's domestic population is euroskeptic, the higher the average number of weakenings. A move from the least euroskeptic Member State, Ireland in 1998 with just 3 percent of the population euroskeptic, to the most euroskeptic Member State, Sweden in 2003 at 59 percent of the population, increases the probability of a weakening by .48. We also check whether or not these estimated effects vary by different country characteristics, as estimates of variables in logit models depend on values of other input variables. In Figure 4, we show the estimated positive relationship between euroskepticism and the probability of weakening the Commission's recommendations for large and small countries. The x-axis, which measures euroskepticism, is restricted to the interval where we actually observe data.

Large countries (solid-line) have an estimated higher probability of weakening than small countries (dotted-line) and high euroskeptic countries are also more likely to weaken than less euroskeptic countries.

Other weak or non-results are also of interest. There is weaker evidence that euro “out” countries are less likely to weaken the Commission’s suggestions—the variable is not statistically significant at either the $p < .05$ or $p < .10$ level in the random effects specification. The domestic institutional variables partisanship and elections do not seem to matter in either specification, and where a Member State is in its economic cycle also does not have a substantive impact.⁴⁷ Finally, we also find evidence of a positive relationship between higher average values of weakening and additional weakenings; this provides some indication that Member States may engage in log-rolling, where a higher number of changes in texts is associated with higher editing.

[Insert Table 1 About Here]

[Insert Figure 4 About Here]

In addition to Member State macro-characteristics, we are also interested in whether or not domestic institutions matter, especially whether domestic elections and the partisanship of the government helps predict rule bending. As we discussed in the theory section, there are some reasons to be skeptical. Coalition politics in particular make the partisanship explanation unlikely in the EU context, however, we test for this explicitly. We find a slight negative relationship between partisanship of the Prime Minister and the probability of weakening, though the relationship is not significant by traditional cutoff points. Upon further inspection, we also find no substantive relationship between partisanship and rule bending. We also run the same model again only this

⁴⁷We considered several alternatives here to get at the underlining concepts. For economic performance, the timing could be relevant, and we checked the performance in the previous and well as current year. We also considered changes instead of levels; if a Member States is experiencing a marked economic slowdown, it may expect a weakened Commission text because of economic factors. Similarly, we examined different lag structures of when an election takes place. None of these alternative specifications was statistically significant or affected the substantive results.

time using the partisanship of the Finance Minister. Here we find slightly stronger negative result; more right-leaning finance ministers are less likely to weaken the Commission's report, however, the results are also insignificant and weak. Finally, we also include measures for both current year and forward looking elections. Again, all estimates are substantively and statistically insignificant. Therefore, one of our significant findings is that we find little evidence that domestic institutional variables matter for explaining rule bending. Furthermore, we think that this null findings is one of the important empirical contributions we make in this paper.

Finally, we also examine the effects of Member State log-rolling. Here the argument is that it is easier for Member States to bend the rules when other Member States are also bending the rules. Because we use the yearly trend in the number of weakenings, the estimate for log-rolling in the fixed effect model is highly correlated with the year fixed effects intercepts. In order to model directly the aggregate time dependencies from the substantively interesting time dependencies, we estimate log-rolling using the random effects specification.⁴⁸ Here we find an estimated higher probability of weakening as the average number of weakening by Member States increases across the sample.

Our next step is to examine whether or not the results are similar when we examine "edit intensity" as rule bending. We run the same analysis as before, with one model with year fixed effects and a second with country and year random effects, but, because the edit distance measure is continuous, we use a linear model. Once again, we find that large states are more likely to have more editing than smaller states. A one-unit increase in size, from small to large Member State, is positively associated with approximately additional 300 edits. Furthermore, as before, being an "out" Member State is negatively associated with editing, though again, the sample size makes the uncertainty intervals around the estimate large here as well – "out" countries are estimated to make on average 100 or so less edits than "in" countries.

⁴⁸see [Clark and Linzer \(forthcoming\)](#) and [Bell and Jones \(2014\)](#). One benefit of the random effects specification is that we can model the intercepts varying among crossed random effects (year, country).

Unlike our previous findings, however, we do not find evidence that euroskeptic countries are more likely to edit intensively. This finding suggests that there is something unique to the strategy of weakening and that euroskeptic countries are more likely to weaken but not necessarily edit intensively. Perhaps even more importantly, like the above findings, we find no systematic evidence that domestic institutional variables matter for rule bending. For example, we find no evidence that elections matter for bending the rules. Finally, the estimate for log-rolling is also positive and significant. As above, these findings imply that a one unit increase in the average intensity of editing generates an additional one edit. This result suggest that Member State behavior depends on what other Member States are doing as well. Given that Member States vote in the Council on each other's Commission-drafted assessments, this empirical finding is consistent with log-rolling within the Council.

[Insert Table 2 About Here]

In sum, the statistical analysis suggests that large Member States make on average a higher number of edits and are more likely to weaken the European Commission's texts about them. In addition, "out" countries are less likely to change the Commission's text and make fewer edits. Member States with relatively more euroskeptic populations are also more likely to weaken text but interestingly, we do not find any evidence that they systematically edit more than their euro-positive counterparts. This finding conforms with the theory that euroskeptic Member States pay lower audience cost at home, which allows them to more easily weaken the EU's rules. It does not, however, explain why they do not try to change the textual information more frequently. Also important for having confidence in our results is that the substantive interpretation of both the logit and the linear probability models produce similar findings across model specifications.

4.4 The Euro Crisis

We also check for differences in the estimated effects before and after the euro crisis by including a dummy variable for the crisis period (2008-2011). In these tests, we find that Member States are on average less likely to make amendments, either weakening or textual changes, though the effects are insignificant at traditional cutoff levels. The fact that the coefficient is negative conforms to the decline in the measures of the dependent variables shown in Figure 2.

Another issue may have to do with varying coefficients for the euro crisis and pre-euro crisis period. As our dataset indicates, the number of edits first increases over the period and then declines. One way to think about this is to consider a simple interaction model. In this case there are two regimes, where the first set of estimated coefficients reports those periods where Member States actively try to edit and weaken the Commission's evaluations, and the second estimated coefficients reports when Member States do not actively try to edit and weaken the Commission's evaluations. We examine this by interacting the crisis variable with the log-rolling variable. This gives us a separate effect of the average amounts of editing conditional on whether or not we are in the pre-crisis or crisis period. Unfortunately, due to data limitations, we cannot evaluate the estimated coefficients after the reforms in 2011. When we account for possible variation in the coefficients in this simple way, the variables for political power and euroskepticism remain stable and significant predictors for weakening. In addition, the crisis year estimate has a negative, though insignificant, estimated relationship with weakening as before. The log-rolling estimate continues to have a positive and significant effect, suggesting that when the average number of edits increases, weakening increases. This is consistent with our previous findings on log-rolling. The estimate for the interaction term of crisis and the spline is positive and insignificant; this suggests that the results hold across the sample broadly.

We also run the same model using the alternative dependent variable, edit intensity. Our findings are similar. Political power remains a positive and significant predictor of edit intensity. The log-rolling variable, measuring the trend number of edits, is also positively associated with edit

intensity, again, suggesting that log-rolling may be at work. The crisis variable is negative though again insignificant by traditional cutoff values. Finally, the interaction between the crisis years and the trend number of edits is insignificant.

In sum, the model we present here seems to hold in the initial years of the crisis.

[Insert Table 3 About Here]

4.5 Robustness Checks

This section provides evidence that the estimates reported above are stable in different model specifications. First, instead of measuring power as all countries with more than 10 votes on the Council, we create an indicator variable for only France and Germany and rerun the results.⁴⁹ We find similar results: political power, in this case measured only by France and Germany, is a strong and stable predictor of textual weakening. In addition, the variable for euroskepticism remains positive and significant. In two other specifications, we change the timing for the euroskepticism variable to include the current year's level of expressed euroskepticism. As we mentioned above, there is some worry that the current year's level may be endogenously related to weakening, however, we use the current year as a robustness check to confirm that the results are stable. Like the other models, we find that euroskepticism is a strong predictor of weakening whether we use the current or forward specification. Furthermore, this remains true whether we specify large country power using only France and Germany, or the larger sample of countries with voting power above 10.

When we run the results using the dependent variable for editing, however, using only the France and Germany specification for power, the political power variable is positive but no longer significant at traditional cutoff points. Thus, there is some evidence that it is not just the behavior of France and Germany, but the larger sample of countries with voting power in the Council, that

⁴⁹We also include the actual votes for all member states under the Treaty of Nice as well as the number of votes needed under Nice to block a Commission proposal for a given Member State. The results are substantively similar to the specification with the dummy variable for the four Member States with more than 10 votes.

matters. Furthermore, we also test the effects of euroskepticism in the current and forward versions of the variables on text editing. Both variables are large and positive, though insignificant using traditional cutoffs, again, consistent with the findings we report in Table 2.

These findings are reported in Table 4.

Instead of being limited to only the EU-15, we also ran the results using an expanded sample of all EU-27 countries. The results, presented in Table 5, are consistent with the results from the restricted sample above. Here we find that countries that are powerful and countries that are more euroskeptic are also more likely to weaken and edit more intensively. While we think that the results from the expanded sample are helpful in bolstering our claims, we still believe that the analysis is most appropriate for the smaller sample of countries. This is because these Member States, as founding members, are states where rule bending behavior should matter most. Their rule bending behavior is also the most interesting as, especially for the “in” countries, it begs the question, why bend the rules after you made the rules? The evidence that we present suggests that not all countries can (and do) bend the rules: the larger, more powerful countries are mostly to blame for weakening the “preventative” arm of the European Commission.

In summary, the results from our robustness analyses are consistent with the evidence that we provide in the previous section. Political power, especially when measured as a Member State with the largest number of votes on the Council, is positively associated with weakening and editing. Also, the level of euroskepticism in the domestic population matters, especially in predicting whether or not Member States will weaken the Commission’s text. What is perhaps most important is what we do not find: in none of the specifications do the domestic institutional variables matter. Therefore, our main empirical finding is that domestic variables such as elections and the partisans hue of the government are not systematically related to observed textual weakening of the Commission’s evaluations, nor do elections or partisanship predict textual editing. Thus, there is little evidence that domestic explanations help to explain rule bending, at least in this context.

5 Conclusion

This paper explores one mechanism at the European level that was meant to prevent a debt-fueled crisis in the Eurozone, namely the Stability and Growth Pact (SGP). We examine in particular whether Member States were able to undermine the “watchdog” function of the European Commission in the Council of Ministers, which is behavior we identify as “rule bending.” We test for the determinants of SGP rule bending using human coding and machine based textual analysis of Commission texts and Council edits. Our sample includes the EU-15, from 1998-1999 to 2012-2013, and our results are also robust to a sample using all EU-27 countries. Unlike in the realm of trade where flexibility has shown to provide real benefits, flexibility in the SGP had adverse consequences. Flexibility did not provide a forum for Member States to make deep commitments. Instead, the ability for large and euroskeptic Member States to amend recommendations generated incentives for Member States to pursue uncooperative strategies. The consequence was that Member State behavior systematically eroded the power of the oversight mechanism, and that this behavior may have contributed to the eurocrisis because Member States did not need to respond to more critical Commission assessments. These, in turn, were a pre-condition for any action under the “corrective” arm of the Pact.

One important caveat is that while domestic population preferences seem to influence the willingness of governments to bend the rules, upcoming elections and the partisan hue of the government do not. These results, of course, may be dependent on the time frame we examine, where there seemed to be little partisan disagreement about the desirability of small budget deficits. The increasingly contentious debates about the limits of “austerity” packages suggests that partisan differences may become more relevant in the future.

Looking forward, one question is whether changes in the framework at the end of 2011, be it through the “Six Pack,” “Two Pack,” or “Fiscal Compact” will change the patterns we observe in our study. Take the two main findings in this paper. In terms of euroskepticism, one goal of the Eu-

ropean Semester is to increase the involvement of national parliaments as “watchdogs.” Moreover, as European election results in May 2014 suggest, European populations are even more euroskeptic than before. This may embolden Member State governments to fight what the Commission says about them and to work harder to change judgments they find unacceptable. Furthermore, there is additional evidence that the level of euroskepticism of a domestic population affects the interaction of domestic institutions on the European stage. One recent study⁵⁰ finds that more euroskeptic parliaments are also the ones that are engaged in checking the Commission recommendations. While the actors in this study are parliaments and not the governments themselves, such evidence suggests that greater euroskepticism may continue to affect the way that national institutions interact with the institutions of the European Union.

When it comes to large Member States, the relationship that we find in this paper might also change. We find evidence that a large state is more likely to change the Commission’s document, but the rules in place through 2011 now require a qualified majority vote to pass the Council’s recommendation. Thereafter, it now takes a qualified majority to block a Commission text. If the formal rules are all that matter, this change in the rules should return some power back to the Commission. Soberly, however, as Stone (2011) documents, large states often have outsized influence in international organizations especially during extraordinary times. Germany has clearly played the type of role Stone describes since the beginning of the euro crisis, and one would anticipate that it would continue to do so now. If we combine more euroskeptic populations with German insistence that the rules now be respected, we would expect increasing tensions on the application of this framework.

Finally, we also show how the use of human coding and computer textual analysis can be fruitfully employed to measure and analyze important aspect of rule bending in international organizations. This is important as scholars often pay attention to whether countries comply with the rules without first uncovering whether and which countries bend the rules too. Equipped with

⁵⁰[Hallerberg, Marzinotto and Wolff \(2014\)](#)

new tools that allow for the computation of textual changes, we show how these tools can be used to contrive of a latent dimension, “rule bending” which can be used as a dependent variable for statistical analysis. Models similar to the one explored here could also be extended to account for deliberation and bargaining in committees, such as in legislative or central bank committees, or as important information that tracks variation in actors’ bargaining positions across space and over time, and thus could be used to analyze textual changes in peace agreements, trade agreements, or multilateral lending negotiations.

Table 1: Explaining Weakening of Commission Recommendations (Logit Analysis)

Dependent Variable: Weakening	Year FE	Year, Country RE
$LargeState_t$	0.81** (0.38)	0.84* (0.48)
$OutputGap_t$	-0.06 (0.14)	0.03 (0.09)
$EuroOut_t$	-0.43 (0.47)	-0.42 (0.57)
$Elections_{t+1}$	-0.35 (0.57)	-0.22 (0.56)
$Euroskepticism_{t+1}$	5.20** (3.39)	5.37** (1.78)
$Logrolling_t$	-17.72** (4.69)	5.74** (1.19)
Intercept	—	-4.48 (0.84)
Number of Observations	209	209
** p < 0.05 * p < 0.10 (Two-tailed.)		

Table 2: Explaining Editing of Commission Recommendations (OLS Analysis)

Dependent Variable: Edit Distance	Year FE	Year, Country RE
$LargeState_t$	310** (97)	308** (117)
$OutputGap_t$	-44 (32)	-22 (19)
$EuroOut_t$	-126 (100)	-114 (123)
$Elections_{t+1}$	-3 (129)	-7 (119)
$Euroskepticism_{t+1}$	-221 (390)	-237 (434)
$Logrolling_t$	0.92** (0.20)	1.05** (0.09)
Intercept	—	-52 (168)
Number of Observations	177	177
** p < 0.05 * p < 0.10 (Two-tailed.)		

Table 3: Effects of Financial Crisis on Weakening and Editing

Dependent Variable:	Weakening RE	Editing RE
<i>LargeState_t</i>	0.84* (0.48)	307** (117)
<i>OutputGap_t</i>	0.01 (0.01)	-23 (20)
<i>EuroOut_t</i>	-0.44 (0.57)	-115 (123)
<i>Elections_{t+1}</i>	-0.25 (0.57)	-8 (119)
<i>Euroskepticism_{t+1}</i>	5.37** (2.00)	-235 435
<i>Logrolling_t</i>	5.68** (1.20)	1.04** (0.95)
<i>EuroCrisis_t</i>	-0.23 (0.49)	-30 (104)
Intercept	-4.37** (0.86)	-27 (190)
Number of Observations	209	177
** p < 0.05 * p < 0.10 (Two-tailed.)		

Table 4: Robustness Check: EU-15

	France and Germany		Coming Election		Excessive Debts		Partisanship of PM and FM		Logrolling during Crisis	
	Weakening	Editing	Weakening	Editing	Weakening	Editing	Weakening	Editing	Weakening	Editing
<i>FranceGermany_t</i>	1.20** (0.58)	253 (172)								
<i>LargeState_t</i>			0.84* (0.48)	298** (120)	0.91* (0.50)	230* (118)	1.07** (0.53)	263* (137)	0.83* (0.48)	306** (118)
<i>OutputGap_t</i>	0.03 (0.09)	-13 (19)	0.04 (0.09)	-17 (18)			0.01 (0.10)	-25 (19)	0.01 (0.10)	-20 (20)
<i>ExcessDeficit_t</i>					-0.07 (0.45)	195** (90)				
<i>EuroOut_t</i>	-0.08 (0.56)	-56 (146)	-0.42 (0.57)	-115 (126)	-0.48 (0.60)	-48 (127)	-0.49 (0.62)	-98 (139)	-0.44 (0.57)	-113 (123)
<i>Euroskepticism_{t+1}</i>	5.42** (1.92)	-68 (458)	5.48** (2.00)	-238 (437)	5.29** (1.95)	-305 (414)	5.46** (2.20)	-370 (484)	5.44** (2.01)	-238 (435)
<i>Election_t</i>	-0.31 (0.56)	-15 (120)							-0.31 (0.56)	-14 (120)
<i>Election_{t+1}</i>			0.23 (0.56)	-218* (116)	0.28 (0.56)	-206* (116)	0.42 (0.61)	-241* (124)		
<i>Logrolling_t</i>	5.70** (1.18)		5.68** (1.17)		5.78** (1.20)		5.94** (1.27)		5.68** (1.19)	
<i>Logrolling_t</i>		1.04** (0.09)		1.04** (0.08)		1.06** (0.08)		1.05** (0.09)		1.02** (0.09)
<i>Crisis_t</i>									-0.18 (0.45)	-48 (104)
Intercept	-4.48** (0.83)	-46 (178)	-4.65** (0.83)	26 (167)	-4.63** (0.82)	-39 (167)	-3.68** (1.16)	80 (269)	-4.43** (0.87)	7 (188)
Number of Observations	209	177	209	177	207	175	193	161	209	177

**p < .05 *p < .1 (Two-tailed)

Table 5: Robustness Check: EU-27

	EU-27	
	Weakening	Editing
$LargeState_t$	1.00** (0.48)	176* (101)
$OutputGap_t$	0.13** (0.06)	−5 (10)
$OutState_t$	0.18 (0.43)	−115 (78)
$Euroskepticism_{t+1}$	5.08** (1.68)	363 (321)
$Election_t$	−0.18 (0.50)	−148 (97)
$Logrolling_t$	6.14** (1.12)	
$Logrolling_t$		1.02** (0.07)
Intercept	−4.85** (0.79)	−123 (140)
Number of Observations	279	217

** $p < .05$ * $p < .1$ (Two-tailed)

Example of Commission and Council Assessments of Stability and Convergence Programs

Examples	Country and Year	Commission Assessment	Council Assessment
No Change	Italy (2002-2003)	8/1/03: "Given Italy's high debt, large primary surpluses will be required for many years."	21/1/03: A) "Given Italy's high debt, primary surpluses on the order of 5% of GDP will have to be maintained for many years."
Weakening	France (2010-2011)	6/7/11: Based on the assessment of the updated stability programme...the Council is of the opinion that the macroeconomic scenario underpinning the budgetary projections is too favourableThe aim of the Renewal of Social Dialogue Act and of the 2008 Labour Market Modernisation Act was to modernise social dialogue and to address the issue of labour market dualism. However, the reform did not tackle contractual segmentation and generally fell short of the measures needed to resolve the dualism of the French labour market.	7/12/11: "Based on the assessment of the updated stability programme...the Council is of the opinion that the macroeconomic scenario underpinning the budgetary projections is C 213/8 Official Journal of the European Union 20.7.2011 EN.....The aim of the Renewal of Social Dialogue Act and of the 2008 Labour Market Modernisation Act was to modernise social dialogue and to address the issue of labour market dualism. The latter Act notably introduced a new procedure for terminating permanent contracts by mutual agreement (rupture conventionnelle) which is now increasingly used. "
Weakening	United Kingdom (2004-2005)	16/2/05: A) "Deficit projections to 2006/2007 have been revised upward relative to the previous update despite an essentially unchanged macroeconomic outlook , while..." B) Paragraph concerning possibility of crossing 3% deficit rule (see Council cell) C) "For 2004/05, evidence of significant progress remains unconfirmed in outturn data and there is a high degree of uncertainty over both expenditures and revenues, leaving the risk noted above of a deficit higher than 3% of GDP."	A) "Relative to the 2003 update, deficit projections for both this financial year and the next have been revised upwards; beyond..." B) In a paragraph warning of UK's potential to break 3% rule, Council adds the following sentence: "However, given the decreased volatility of the UK economy in recent years and the cautious approach on the growth assumptions underlying the public finance projections, the risk appears small in the outer years." C) "There is a clear risk that the budgetary outcome could be worse than projected in the programme in the short term."

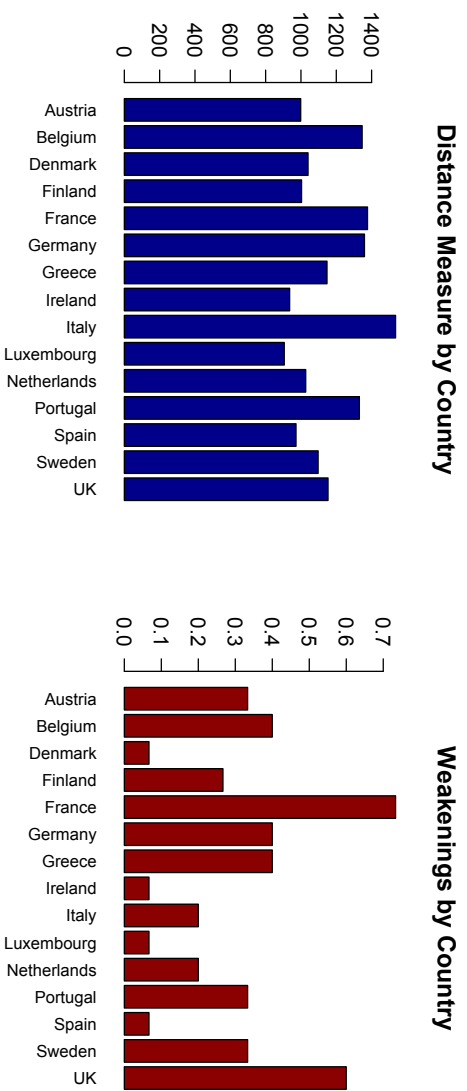


Figure 2:

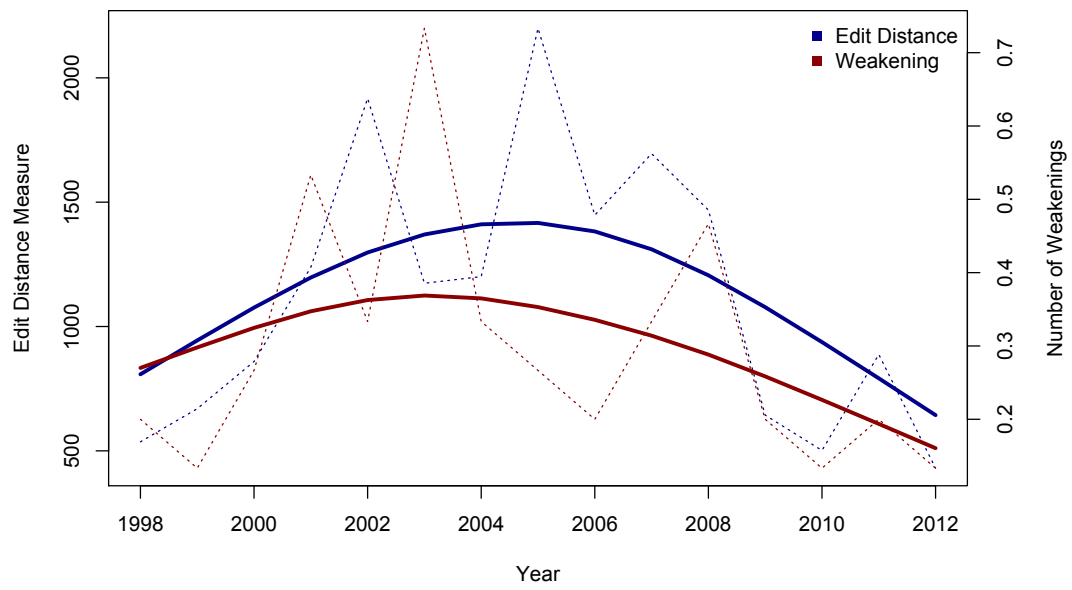
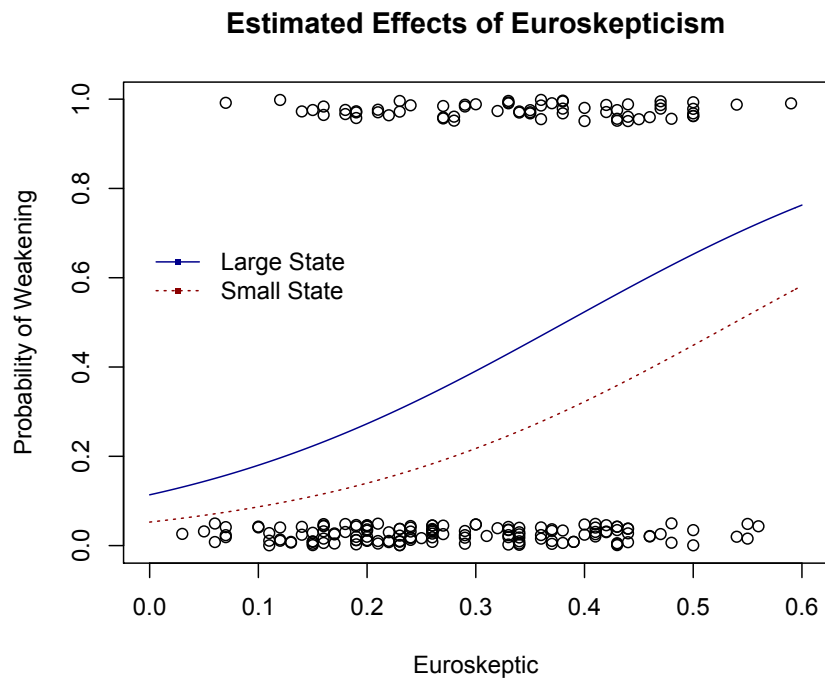


Figure 3:

Figure 4:



6 Appendix A

We construct two dimensions of rule-bending using both human coding and machine learning techniques. The human coding follows procedures discussed above. Documents are compared for substantive changes to the textual draft changes between the Commission's suggestions and the Council's amendments. This approach measures the direction of the recommendation, but does not measure the intensity or the magnitude of the changes.

In order to examine the intensity, we turn to machine learning techniques from computational social sciences. The machine learning follows the procedures summarized above. We first compile a list of textual changes using textual comparison software. We then use an add on package to extract the reported changes into a metadata file. Using two user made python scripts, we then extract the old text and new text and, running a levenshtein distance measure using the `nltk.metrics` package from NLTK, we then compute the metric for each edit. Finally, we then sum the total number of edits by country year and use this as a measure of country-year edit intensity. Replication material for the statistical analysis is available at [Rule Bending Replication Material](#).

References

- Begg, Iain, M Buti, Hendrik Enderlein, L Pench, Waltraud Schelkle and M. Weale. 2004. “Reforming Fiscal Policy Co-ordination under EMU: What Should Become of the Stability and Growth Pact?” *Journal of Common Market Studies* 42(5):1023–1059.
- Bell, Andrew and Kelvyn Jones. 2014. “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data.” *Political Science Research and Methods* pp. 1–21.
- Benoit, Kenneth and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.
- Carrubba, Clifford. 2005. “Courts and Compliance in International Regulatory Regimes.” *The Journal of Politics* 67(3):669–689.
- Chapman, Terry, Li Xin, Songying Fang and Randy Stone. Working Paper. “Mixed Signals: Crisis Lending and Capital Markets.” Working Paper.
- Chaudoin, Stephen. 2014. “Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions.” *International Organization* 68:1–22.
- Clark, Tom and Drew Linzer. forthcoming. “Should I Use Fixed or Random Effects?” *Political Science Research and Methods* .
- Copelovitch, Mark. 2010. “Master or Servant? Common Agency, Preference Heterogeneity, and the Political Economy of IMF Lending.” *International Studies Quarterly* 54(1):49–77.
- De Haan, Jakob, Helge Berger and David-Jan Jansen. 2004. “Why has the Stability and Growth Pact Failed?” *International Finance* 7(2):235–60.
- Dreher, Axel and Roland Vaubel. 2004. “The Causes and Consequences of IMF Conditionality.” *Emerging Markets Finance & Trade* 40(3):26–54.

- Finlayson, Jock A. and Mark W. Zacher. 1981. "The GATT and the Regulation of Trade Barriers: Regime Dynamic and Functions." *International Organizations* 35(4):561–602.
- Fischer, Jonas, Lars Jonung and Martin Larch. 2006. "101 Proposals to reform the Stability and Growth Pact. Why so many? A Survey." *Economic Papers* 267:3–48.
- Franzese, Robert J. 2002. "Electoral and Partisan Cycles in Economic Policies and Outcomes. , 2002: Vol. 5: 369-421." *Annual Review of Political Science* 5:369–421.
- Grieco, Joseph M., Christopher F. Gelpi and T. Camber Warren. 2009. "When Preferences and Commitments Collide: The Effect of Relative Partisan Shifts on International Treaty Compliance." *International Organization* 63(2):341–355.
- Hallerberg, Mark. 2004. *Domestic Budgets in a United Europe: Fiscal Governance from the End of Bretton Woods to EMU*. Ithaca, NY: Cornell University Press.
- Hallerberg, Mark, Benedicta Marzinotto and Guntram Wolff. 2014. "The Europeanization of Euroskeptical Parliaments? The Changing Role of National Parliaments Under the European Semester." Manuscript.
- Hallerberg, Mark, Rolf Strauch and Jürgen von Hagen. 2009. *Fiscal Governance: Evidence from Europe*. Cambridge University Press.
- Heipertz, Martin and Amy Verdun. 2010. *Ruling Europe: The Politics of the Stability and Growth Pact*. Cambridge University Press.
- Jensen, Nathan. Working Paper. "Political Regimes and Political Risk: Democratic Institutions and Expropriation Risk for Multinational Investors." Working Paper.
- Koehler, Sebastian and Thomas König. forthcoming. "Fiscal Governance in the Eurozone: How effective is the Stability and Growth Pact limiting public debt?" *Political Science Research and Methods* .

- König, Thomas and Brooke Luetgert. 2009. "Troubles with Transposition? Explaining Trends in Member-State Notification and the Delayed Transposition of EU Directives." *British Journal of Political Science* 39:163–194.
- König, Thomas and Lars Mäder. 2013. "Non-conformable, partial and conformable transposition: A competing risk analysis of the transposition process of directives in the EU15." *European Union Politics*. 14(1):46–69.
- Konrad, Kai and Holger Zschäpitz. 2010. *Schulden ohne Sühne? Warum der Absturz der Staatsfinanzen uns alle trifft*. C.H. Beck.
- Kucik, Jeffrey and Eric Reinhardt. 2008. "Does Flexibility Promote Cooperation? An Application to the Global Trade Regime." *International Organization* 62:477–505.
- Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. 2002. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organizations* 56(3):477–513.
- Morris, Richard, Hedwig Ongena and Ludger Schuknecht. 2006. "The Reform and Implementation of the Stability and Growth Pact." European Central Bank Occasional Paper No. 47, June.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.
- Richard, Stephanie J. and Teri L. Caraway. 2014. "International Negotiations in the Shadow of Elections." *International Organization* 68(3):701–720.
- Schelkle, Waltraud. 2005. "The Political Economy of Fiscal Policy Coordination in EMU: From Disciplinarian Device to Insurance Arrangement." *Journal of Common Market Studies*. 43(2):371–391.

- Schuknecht, Ludger. 2004. "EU Fiscal Rules: Issues and Lessons from Political Economy." European Central Bank Working Paper Number.
- Schuknecht, Ludger, Philippe Moutot, Philipp Rother and Jürgen Stark. 2011. "The Stability and Growth Pact: Crisis and Reform." European Central Bank Working Paper Number 129.
- Sörensen, Kenneth. 2007. "Distance measures based on the edit distance for permutation-type representations." *Journal of Heuristics* 13(1):35–47.
- Stone, Randall. 2011. *Controlling Institutions: International Organizations and the Global Economy*. Cambridge: Cambridge University Press.
- Thacker, Strom. 1999. "The High Politics of IMF Lending." *World Politics* 52(1):38–75.
- Thomson, Robert, Rene Torenvlied and Javier Arregio. 2007. "The Paradox of Compliance: Infringements and Delays in Transposing European Union Directives." *British Journal of Political Science* 37:685–709.
- Tomz, Michael. 2006. "The Effects of International Agreements on Foreign Policy Preferences." Working Paper.
- Tsebelis, George and Xenophon Yataganas. 2002. "Veto Players and Decision-making in the EU After Nice: Policy Stability and Bureaucratic/Judicial Discretion." *Journal of Common Market Studies* 40(2):283–307.