Assessing the Effectiveness of Non-Majoritarian Fiscal Surveillance A Time-Series Analysis

Inaugural dissertation
submitted in partial fulfillment of the requirements for the degree
Doctor of Social Sciences in the
Graduate School of Economic and Social Sciences
at the University of Mannheim

vorgelegt von

Sonja Pohle
aus Mannheim

Mannheim, 2019

Dekan: Prof. Dr. Michael Diehl

Gutachter:
Prof. Dr. Thomas König, Universität Mannheim
Prof. Nicole Rae Baerg, Ph.D., University of Essex
Prof. Dr. Richard Traunmüller, Universität Mannheim

Tag der mündlichen Prüfung: 15. Februar 2019

Abstract

Little research has examined why the European Commission is so hesitant in opening an Excessive Deficit Procedure (EDP) against EU member states with deficit and debt levels that violate the Stability and Growth Pact (SGP). In order to explain this, I argue that the European Commission can be classified as a non-majoritarian institution (NMI) that lacks direct legitimization via elections and complete information about the nature of non-compliance. When sanctioning elected governments, it therefore has to rely on voters' support for compliance with the SGP and a sufficient amount of transparency about the country's economic situation. In order to test this argument empirically, I present a Bayesian multi-level probit model as a new method for analyzing small N panel-heteroskedastic and contemporaneously correlated binary time-series cross-sectional (BTSCS) data and test its compared with existing methods. I find evidence for the suitability of the Bayesian multilevel probit model for modeling BTSCS characteristics, but not for the effect of strong public support and transparency on the Commission's EDP initiations. My contributions are therefore theoretical, empirical and methodological. Theoretically, I combine arguments from European Union research and judicial politics on NMI decision-making to explain the sanctioning of SGP violations by the European Commission. Empirically, I conduct the first quantitative analysis to test these arguments and assess the determinants of the European Commission's sanctioning behavior. Methodologically, I provide researchers with an alternative technique for BTSCS data analysis.

Contents

List of Figures

List of Tables

1 Introduction

Fiscal discipline is increasingly monitored by supranational and/or national independent agencies (Bova et al. 2015). These monitoring institutions – like central banks and constitutional courts – can be referred to as NMIs, namely "[..] governmental entities that [...] possess and exercise some grant of specialised public authority, separate from that of other institutions, but [...] are neither directly elected by the people, nor directly managed by elected officials" (Thatcher and Sweet 2002, 2).

Surprisingly, despite fiscal surveillance of NMIs, rising levels of budget deficits and public debt around the world can be observed (Reinhart and Rogoff 2010; Reinhart et al. 2012; Roubini and Sachs 1989; Alesina et al. 1999; Bova et al. 2015; Grilli et al. 1991).

The most prominent case of the failure of non–majoritarian monitoring of fiscal policy is the "Stability and Growth Pact" (SGP) of the Economic and Monetary Union (EMU) of the European Union (EU) (Buti and Carnot 2012, 899–900). The fiscal surveillance framework outlined by the pact includes the so–called "corrective arm", where the European Commission is responsible for sanctioning a government that violates the pact's deficit and debt reference values by opening an EDP in order to enforce a balanced budget (Buti and Carnot 2012, 900; European Commission 2017d; Köhler et al. 2018, 1–2; Baerg and Hallerberg 2016, 970–973). In the context of the SGP, the European Commission functions as the EU's non–majoritarian central monitoring agency for fiscal policy and it serves as a "watchdog" for compliance with the fiscal rules of the Maastricht treaty (Baerg and Hallerberg 2016, 968–972; Köhler et al. 2018, 9; König and Mäder 2014; van der Veer and Haverland 2018, 2).

The conditions when the Commission should initiate an EDP are clearly defined: the pact prohibits the breach of the reference values of a 3% deficit of GDP and a debt burden of 60 % of GDP (Baerg and Hallerberg 2016, 968–972; Köhler et al. 2018, 9; König and Mäder 2014; van der Veer and Haverland 2018, 2). However, despite these clear criteria, the effectiveness of the SGP in general and the sanctioning behavior of the Commission in particular has been called into question in the face of highly indebted Eurozone member states and the recent European sovereign debt crisis (Baerg and Hallerberg 2016; Fabbrini 2013; King and Bouvet 2011; Howarth 2004; Köhler and König 2015; van der Veer and Haverland 2018). This raises the question of how the lenient sanctioning of the European Commission and consequently the ineffectiveness of the SGP can be explained.

Both economists and political scientists have investigated the determinants of deficit and debt accumulation and institutional solutions that are able to prevent this problem (Alesina and Perotti 1995; Alesina et al. 1999; Annett 2001; Barro 1979; Buchanan and Wagner 1977; Cukierman and Meltzer 1989; De Haan and Sturm 1994; Grilli et al. 1991; Hallerberg and Von Hagen 1999; Kontopoulos and Perotti 1999; Rodden 2002; Roubini and Sachs 1989; Tabellini 1991; Alesina and Tabellini 1990; Persson and Svensson 1989; Velasco 2000; von Hagen and Harden 1995; von Hagen 2002; von Hagen 2006; Wibbels 2000; Weingast et al. 1981).

However, most of this literature focuses on domestic factors and does not consider supranational factors such as the monitoring of fiscal policy by the Commission under the SGP. In turn, the existing literature on the SGP mainly concentrates on limitations in the institutional design of the pact that further a violation of the SGP's deficit and debt criteria and consequently limit its effectiveness (Annett 2006; Köhler and König 2015; Irlenbusch et al. 2003; von Hagen and Wolff 2006). Accordingly, this research ignores the role of the European Commission to sanction rule breaking in a timely manner to maintain the functioning of the SGP. Regarding the Commission, the occurrence of the European sovereign debt crisis showed that the economic criteria as outlined in the pact do not seem to be exclusively decisive for the Commission initiating an EDP.

The literature focusing on Commission behavior identifies the influence of powerful member states on the sanctioning procedure at the EU level as an important determinant for the Commission's hesitant sanctioning (Baerg and Hallerberg 2016; Eichengreen 2005; De Haan et al. 2004; Heipertz and Verdun 2010). The problem is that this explanation exclusively focuses at the European level and ignores determinants at the member state level.

In contrast to these previous explanations, I argue that the Commission suffers from a legitimacy and a transparency problem which affects its decision to sanction overspending governments. On the one hand, as an NMI, the Commission lacks direct electoral legitimization and therefore needs to consider voters' support for itself when sanctioning elected member state governments (Caldeira and Gibson 1992, 695; Köhler et al. 2018, 11–15;

Stéclebout-Orseau and Hallerberg 2007, 10–22; Thatcher and Sweet 2002, 1–2; Vanberg 2001, 347; Vanberg 2005, 20–24). In order to compensate for this lack of legitimization via direct elections, it relies on voters' support at the member state level: only when the public supports the European Commission and the rules of the SGP does the Commission have sufficient political power to open an EDP against a non–compliant member state (Köhler et al. 2018, 11–15; Vanberg 2001, 347; Vanberg 2005, 19–24).

On the other hand, the Commission acts under uncertainty when sanctioning member state governments since it has no information about whether excessive fiscal policy is justified – for example, in the occurrence of a sudden economic crisis – or, caused by serving voters' consumptive interest with high amounts of public spending (Köhler et al. 2018, 14–19). In order to avoid antagonizing voters via an unjustified EDP initiation – which would endanger the Commission's future legitimacy – it is crucial for the Commission to make the appropriate decision about whether the government's overspending is due to responsiveness (Köhler et al. 2018, 14–15). A possible solution for the Commission's dilemma is increased transparency: from judicial politics, it is known that courts need a transparent *political system* so that the public can detect government non–compliance (Köhler et al. 2018, 14–15; Vanberg 2005, 19–49). However, in contrast to a court, I argue that the Commission is in need of transparency about the country's *economic situation* to identify whether an EDP initiation is appropriate. In this regard, the absence of an economic crisis indicates a responsive government which justifies an EDP initiation without undermining the Commission's legitimacy. Consequently, the Commission will only sanction SGP violations via an EDP if its underlying problems of insufficient legitimacy and transparency are solved.

I test the effect of these two factors on EDP initiations on a data for twenty-five EU–members, all of which violated the pact during the period 2002 to 2017. By comprising repeated observations across different countries, a binary dependent variable and a low number of countries and years, the data for this investigation can be classified as binary time–series cross–sectional (BTSCS) data (Beck et al. 1998; Beck 2001a, 111–273; Beck 2001b, 271; Beck and Katz 2007, 183). Analyzing this data bears a number of methodological challenges, such as dealing with the presence of contemporaneous correlation and panel heteroskedasticity in a small sample, which can lead to incorrect standard errors and even biased estimates in logit or probit models (Beck and Katz 1995, 636; Beck et al. 1998, 1261–1263; Beck 2001a, 118–130; Beck 2001b, 275–288; Shor et al. 2007, 165–172; Yatchew and Griliches 1985, 135–137). Unfortunately, standard methods for BTSCS investigations such as the temporal dummy approach proposed by Beck et al. (1998) or multi–level models including country–level intercepts ignore either panel heteroskedasticity or contemporaneous correlation (Beck et al. 1998, 1261–1283; Beck 2001a; Beck 2001b; Fahrmeir et al. 2013; Gelman and Hill 2006; Hedeker and Gibbons 2006). A promising alternative are Bayesian multi–level models that have been extended by a year–level random intercept, thereby accounting for panel heteroskedasticity and contemporaneous correlation at the same time, while resulting in correct standard errors even for samples with few countries (Pang 2010; Shor et al. 2007; Stegmüller 2013). Since evidence on the performance of this alternative to standard techniques is limited for nonlinear models, I assess the performance of a Bayesian multi–level probit (BMLP BTSCS) model with a country– and year–specific random intercept compared with the performance of five other models used for BTSCS data analysis – including Beck et al.'s (1998) BTSCS standard approach and frequentist multi–level probit models – in a Montel Carlo experiment. I find that the model yields unbiased estimates and correct standard errors even for small sample sizes with $10 \leq N \leq 30$, which makes it an appropriate technique for estimating the effect of public support and transparency on the initiation of EDPs by the European Commission.

The results of the analysis indicate that EDP openings by the European Commission is not unambiguously influenced by both public support and transparency: high levels of both factors do not increase the probability of EDP initiations in case of non–compliance with the SGP in all model specifications and the uncertainty of the results is high. Nevertheless, the results hint at a "democratic deficit" of the Commission, resulting from its lack of legitimacy as one reason for the ineffectiveness of the SGP, as it induces a strategic sanctioning behavior of the Commission (Follesdal and Hix 2006; Majone 1998; Moravcsik 2002).

The contribution of this research is threefold: First, it adds to the body of literature on explaining SGP (in)effectiveness by combining theoretical arguments from EU research and judicial politics on the strategic nature of NMI decision–making. In this regard, I identify the lack of legitimacy and transparency as key determinants of the malfunctioning of non–majoritarian fiscal surveillance in the EU. Second, by testing the effect of public support and transparency on EDP initiations by the Commission in a quantitative analysis, I offer the first study providing an empirical assessment of the determinants of NMI decision–making in the context of the SGP. Third, the research contributes to the methodological literature on BTSCS analysis by addressing the methodological question of which model is most appropriate for small N BTSCS data.

The remainder of this thesis is structured as follows. First, I begin with an overview of the political economy of deficit– and debt–making, as well as the existing national institutional solutions to this problem. I then continue with a more specific literature review on the effectiveness of the SGP and non–majoritarian economic governance within the EMU.

In the second part, I present and discuss the theoretical models that attempt to explain NMI decision–making and combine their predictions to explain the Commission's decision making process when initiating EDPs.

Subsequently, I discuss the methodological challenges of European Union research in the third section. Here, I also describe the research design including the data and the method to be used as well as the operationalization of the variables included in the statistical model. Moreover, I explain the characteristics of BTSCS data and methodological solutions for dealing with them in further detail. Finally, I present the Monte Carlo experiment with the goal of finding the best method for assessing the determinants of EDP initiations by the European Commission.

Having determined the model to be chosen for the analysis, the fifth section includes some descriptive statistics and the results of the analysis. Finally, these results and their implications are summarized and discussed in the conclusion.

2 Why do states overspend?

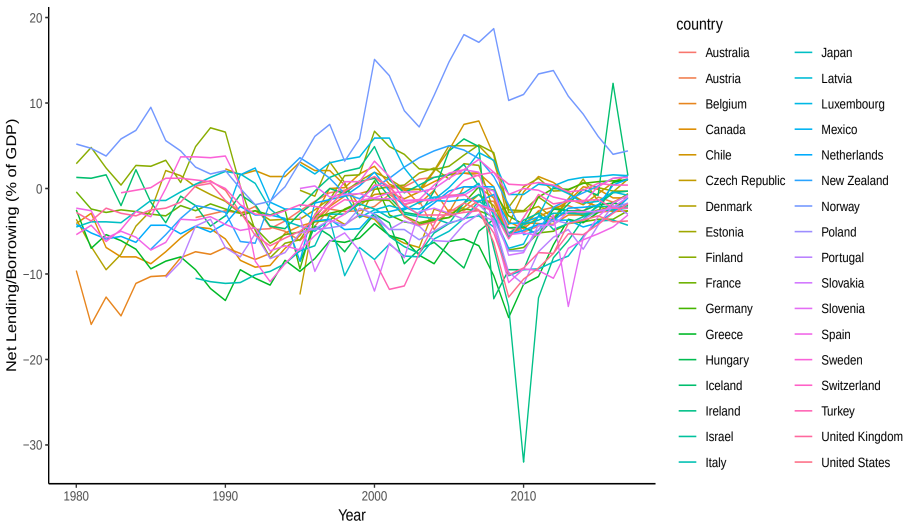


Figure 1: Budget Deficits of OECD Countries since 1980

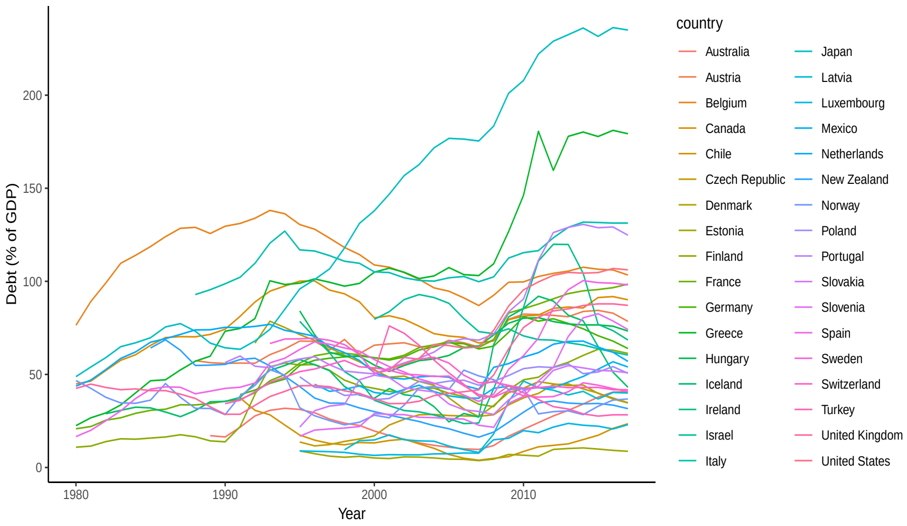


Figure 2: Debt Levels of OECD Countries since 1980

Figures 1 and 2 [1]) show the net lending/borrowing rate and general government gross debt as a percentage of GDP of all OECD countries from 1980 to 2017. The graphs show two interesting developments: First, the net lending/borrowing rate is below zero for most OECD countries, which indicates a budget deficit. Moreover, we also observe an increasing trend for many OECD countries in terms of debt. However, second, there remains considerable variation across these countries: for example, while Norway runs a surplus, Ireland experienced a deep deficit. The same is true for debt levels, with Japan running the highest debt level and Denmark maintaining the lowest amount of debt. High deficits and debt levels thus do not seem to be an exclusively European phenomenon.

The graphs lead to two questions: Why do so many countries accumulate high deficit and debt levels? What explains the variation in deficit and debt levels across countries? Both economists and political scientists have attempted to answer this question with different explanations.

A starting point is the so–called "tax–smoothing" model, which assumes that the government is a unitary actor that acts like a "benevolent social planner" seeking to maximize the utility of the "representative agent", namely the people (Alesina and Perotti 1995, 4–5; Barro 1979; Lucas Jr and Stokey 1983). Since taxes rate should be kept constant and taxes

[1]Figures 3 and 4 have been created using the R–packages `dplyr`, `gdata`, `ggplot2` and `foreign` (R Core Team 2015; Warnes et al. 2017; Wickham 2016; Wickham et al. 2017). Data are obtained from the IMF database (International Monetary Fund 2019b; International Monetary Fund 2019a).

are determined by the intertemporal budget constraint – which postulates that the present value of spending needs to be equal to the present value of spending – budget deficits are used as a “buffer” given that deficits occur when spending is temporarily high (due to e.g. economic shocks or – in former times – war), while surpluses occur when spending is low (Alesina and Perotti 1995, 6–7; Barro 1979, 940–943; Lucas Jr and Stokey 1983, 99). Barro (1979) finds evidence for the tax–smoothing model, given that debt increases as expected during wartime such as the Korean and Vietnam War and recessions (Barro 1979, 960–969).

The tax–smoothing model provides a purely economic explanation for deficit and debt accumulation, but leaves the question of why some countries maintain higher deficit and/or debt levels than others unanswered (Alesina and Perotti 1995, 9–10). While OECD countries are very similar economically as they are all developed countries, they vary in their political characteristics. Thus, scholars have deviated from an exclusively economic approach and moved towards political–economic explanations that relax the unitary actor assumption of the government and focus more on the relation between governments and voters.

The first political–economic approach is models based on the idea of “fiscal illusion” (Alesina and Perotti 1995, 10; Buchanan and Wagner 1977, 133–134). The key argument is that a complex taxation system prompts voters to underestimate the true costs of public goods and services and therefore demand higher levels of them (Buchanan and Wagner 1977, 131–136). At the same time, office–seeking politicians have the opportunity to finance voters’ demand by debt, which further increases both the misperception of the costs of public goods and services and their demand (Buchanan and Wagner 1977, 144). As politicians’ incentive to raise taxes is very low in such a situation, the result are high deficit and debt burdens (Alesina and Perotti 1995, 11;Buchanan and Wagner 1977, 144).

The fiscal illusion approach is related to the literature on “political business cycles”, which predicts pre–electoral expansionary policies followed by recessions (Alesina and Perotti 1995, 13; Nordhaus 1975). Even though the original political business cycle model developed by Nordhaus (1975) focused on monetary policy and its inherent inflation/unemployment trade–off, its logic also applies to fiscal policy given that we will observe budget deficits before an election (Alesina et al. 1993, 16; Alesina and Perotti 1995, 13; Alt and Lassen 2006; Nordhaus 1975). However, the empirical evidence on fiscal political business cycles is mixed and seems to depend on third factors such as the transparency and polarization of the political system (Alesina et al. 1993; Alt and Lassen 2006). Besides the mixed empirical evidence, it is unclear why voters should always be biased in the same way – namely overestimating benefit and underestimating tax burden – and not the other way round: it is also possible that voters discard useful long–term investments that are expected to bring high revenues because they overestimate their costs (Alesina and Perotti 1995, 11). Moreover, this model is also unable to explain cross–country differences in democracies, where in theory all voters should be fiscally biased in the same way (Alesina and Perotti 1995, 12).

While the fiscal illusion approach focuses on the relationship between voters and governments to explain deficit– and debt–making, models of intergenerational redistribution concentrate on the relationship between generations of voters (Alesina and Perotti 1995, 13). This category of models relaxes the principle of "Ricardian equivalence", which states that the distribution of tax burden is independent from debt because changes in the amount of public debt are compensated by changes in private bequests given a sufficient amount of altruism (Alesina and Perotti 1995, 13–14). If this principle no longer holds, it leads to the intergenerational redistribution of debt as the current generation – which is eligible to vote at time t – transfers the debt burden to the future generation who are not yet allowed to vote (Alesina and Perotti 1995, 14).

Cukierman and Meltzer (1989) argue that the Ricardian equivalence is unlikely to hold for all voters as they have different abilities and will in leaving bequests that compensate future tax burdens (Cukierman and Meltzer 1989, 713). Some voters even prefer to run budget deficits to increase their own consumption and therefore leave negative bequests (Cukierman and Meltzer 1989, 713). Consequently, the debt accumulation of a nation reflects the number of people – and the majority of voters – who are leaving negative bequests (Alesina and Perotti 1995, 15; Cukierman and Meltzer 1989, 730).

Intergenerational redistribution is a feasible strategy because future generations will repay the debt issued by former generations, given that they were unable to vote at the time when the debt was issued and repudiating it would harm not only the old but also the wealthy (Tabellini 1991, 336). To avoid this, there will always be a fraction of young voters who are rather in favor of repaying debt, which makes it possible that former generations are able to transfer their debt to subsequent generations (Tabellini 1991, 336).

While these theories are able to explain the reasons for debt accumulation, the question about the reasons for cross–country variation remains unanswered by these theories, as this would imply that intergenerational altruism is stronger in some countries than others (Alesina and Perotti 1995, 15).

Instead of concentrating on the voter level, models that regard debt as a strategic variable focus on the interaction between consecutive governments (Alesina and Perotti 1995, 16). In these models, governments issue debt as a strategic means to restrict the acts of succeeding governments (Alesina and Perotti 1995, 16–22).

Alesina and Tabellini (1990) relax the assumption of a benevolent social planner by Barro (1979), and model the interaction between alternating policymakers with different preferences about the *composition* of public spending resulting from elections (Alesina and Tabellini 1990, 403–412). If each of these policymakers uses debt to strategically constrain their successor, high political polarization leads to an accumulation of debt that is higher than socially optimal (Alesina and Tabellini 1990, 404).

Somewhat relatedly, Persson and Svensson (1989) also focus on alternating policymakers (a conservative government and a liberal government), although in contrast to Alesina and Tabellini (1990) these two policymakers have different preferences about the *level* rather than the composition of public spending (Alesina and Tabellini 1990, 412; Persson and Svensson 1989, 325–340). They find that a conservative government – knowing that it will be replaced by a liberal one – will leave a higher amount of public debt to its successor, thereby constraining the latter's level of spending (Persson and Svensson 1989, 339–341). While the budget deficit of the Reagan administration can be explained by this theory, it also introduces the amount of political polarization and political competition which are able to explain cross–country variation (Alesina and Perotti 1995, 21–22; Persson and Svensson 1989, 326). These two factors play a stronger role in the next class of models of distributional conflicts within social groups and/or coalition partners and models of geographically–dispersed interests (Alesina and Perotti 1995, 5).

In these models, deficit and debt accumulation is a result of intragovernmental (for example, among coalition partners) rather than intergovernmental conflict (Alesina and Perotti 1995, 22). According to Weingast et al. (1981), public decision–making suffers from an inefficiency bias towards large but inefficient projects, namely projects or grants from which only specific constituencies benefit (Weingast et al. 1981, 643). This bias stems from three sources: first, locally–targeted expenditures are only seen as benefits by the respective local constituency (Weingast et al. 1981, 658); second, the existence of separate districts whereby only *one* specific constituency benefit from a project combined with general taxation that distributes the costs of this given projects across *all* constituencies incentivizes policymakers to increase a project to an inefficient size as the link between beneficiaries and revenue sources is weakened (Weingast et al. 1981, 643–658); and finally, the previous two sources of inefficiency remain because they maximize total political net benefits (Weingast et al. 1981, 657–658).

The "common–pool problem of public finances" is related to the model of Weingast et al. (1991) (von Hagen 2002, 263–264; von Hagen 2006, 464–465). The common–pool problem arises in a situation where politicians have access to a general tax fund from which they can spend money on some target group (von Hagen 2006, 465). The main problem of this situation is that the net beneficiaries of this spending are often small for the whole society, as the target group is usually much smaller than the group of taxpayers (von Hagen 2006, 465). As a result, these groups and politicians representing them ask for increasingly more spending on policies that are mainly in their interest, but not in the interest of society as a whole (von Hagen 2006, 465). It is exactly this dynamic via which the common–pool problem leads to high deficit and debt levels (von Hagen 2006, 465).

The common–pool problem's negative effect on deficit and debt levels is intensified by the number of politicians who have access to the tax fund, ideological and/or ethnical and vertical fiscal division, which has been well confirmed empirically (Annett 2001, 561; De Haan and Sturm 1994; Grilli et al. 1991; Kontopoulos and Perotti 1999; Rodden 2002, 675–682; Roubini and Sachs 1989); Velasco 2000; von Hagen 2006, 465;Wibbels 2000, 687–694).

For example, the number of spending ministers ("cabinet size") (Kontopoulos and Perotti 1999, 81–101); coalition size (Kontopoulos and Perotti 1999, 81–101; Roubini and Sachs 1989, 931), proportional representation with a fragmented party system (Grilli et al. 1991, 349–375), frequent government changes (De Haan and Sturm 1994, 157–161; Roubini and Sachs 1989, 904–905) and left–wing partisanship (De Haan and Sturm 1994, 157–161) contribute to higher government spending, deficit and debt levels. Annett (2001) confirms that ethnolinguistic and religious fractionalization increases government consumption, as this is a necessary means to maintain political stability (Annett 2001, 561–589). Federalism where subfederal units are allowed to borrow money and have a high amount of fiscal autonomy have also been found to accumulate higher deficits and have higher debt levels than unitary states where provinces lack these rights (Rodden 2002, 675–682; Wibbels 2000, 687–694).

Since states vary in these political characteristics, models of distributional conflicts within social groups and/or coalition partners and models of geographically–dispersed interests finally provide an explanation for both the rising high levels of deficit and debt and the variation across countries in terms of these two economic variables observed in

Figures1 and 2. However, answering the first two questions leads to a third one that has not been answered by any of the aforementioned models: What institutional solutions that can prevent states from accumulating a high deficit and/or a high debt burden?

3 Institutional Solutions to Prevent Governmental Overspending

While the reasons for overspending have been identified, the question regarding appropriate institutional solutions for these problems remains, which is the focus of the final class of models, models on budgetary institutions (Alesina and Perotti 1995, 5). According to these models, budget institutions have to fulfill two conditions to be effective: they need to be difficult to change and they need to influence the final vote and implementation of the budget (Alesina and Perotti 1995, 32).

To date, the literature identified three institutional solutions that help to mitigate the common–pool problem on deficit and debt accumulation: first, the establishment of *ex–ante fiscal rules*; second, *electoral rules* that further both competition and accountability; and third, *modifications of the procedures of the budget process* (von Hagen 2002, 264–265; von Hagen 2006, 465).

Ex–ante fiscal rules can take the form of balanced budget constraints, numerical debt ceilings, and limits on tax and/or spending growth (von Hagen 2006, 466).

Electoral rules that prevent politicians from extracting high rents are characterized by small district magnitude and plurality rule where individual candidates can be elected (von Hagen 2006, 468). This is because the focus on individual candidates in such systems leads to close monitoring of this candidate's performance, thereby maximizing his/her accountability (von Hagen 2006, 468).

Budgetary decision–making processes also need to be designed in a way that politicians take a more comprehensive view of their budgetary decisions in terms of the costs and benefits to society as a whole – rather than only their constituencies – as highlighted by Weingast et al. (1981) – in order to effectively reduce budget deficits and debt (von Hagen 2006, 470; (Weingast et al. 1981). A budgetary process that is both centralized and transparent achieves this goal: centralization means that all conflicts regarding the budget are solved within this process, while transparency requires comprehensive budget documents and clear attribution of expenditures (von Hagen 2006, 470).

In contrast to the unambiguous concept of transparency, centralization can usually be achieved by either delegation or contracts (von Hagen 2006, 471). Delegation means that one agent – typically the finance minister – sets the rules for the budget and enforces them against (over)spending ministers (von Hagen 2006, 471). Contracts usually refer to the situation in which all members of the executive agree on a fixed set of fiscal targets (von Hagen 2006, 471).

The effectiveness of each of these three measures has been confirmed by empirical studies: hierarchical budgetary processes with fiscal targets, a strong position of the finance minister, limits on amendments in the parliamentary process, item–by–item votes on expenditures, the decision on a fixed budget by the parliament previous to the debate, a highly transparent budget and limited flexibility for spending ministers have been linked with significantly lower deficit and debt ratios (Alesina et al. 1999, 257–265; von Hagen and Harden 1995, 776–777).

There is also evidence that the three institutional solutions on limiting budget deficits are not mutually exclusive. Hallerberg and von Hagen (1999) argue that the electoral system determines the effectiveness of ex ante rules and modifications in the budget process (Hallerberg and Von Hagen 1999, 211). While the delegation approach is can be applied in plurality electoral systems, parties in coalition governments that exist in proportional representation systems will be less likely to delegate extensive monitoring and sanctioning rights to one minister, particularly if he/she is a member of the other coalition party (Hallerberg and Von Hagen 1999, 211). Instead, negotiated fiscal targets – ideally written down in the coalition agreement – are a more appropriate solution for PR systems (Hallerberg and Von Hagen 1999, 211–224). An empirical test of their argument on fifteen EU states supports their argument, as one–party governments more often choose a delegation approach and multiparty governments the agreement on fiscal targets (Hallerberg and Von Hagen 1999, 211–212). Moreover, while both the delegation and contract approaches are effective in limiting budget deficits, the delegation approach is more effective for single–party governments and the contract approach for coalition governments (Hallerberg and Von Hagen 1999, 229–230). This implies that the absence of appropriate institutional solutions for the prevention of budget deficits rather than the electoral system itself is crucial for fiscal performance (Hallerberg and Von Hagen 1999, 212). Although a country can usually hardly change its electoral system, it is nevertheless free to choose the institutional solution that fits its electoral system to prevent overspending.

From the existing literature on institutions that enhance fiscal discipline we have seen that there exist many solutions whose effectiveness seems to be empirically confirmed. However, this literature exclusively focuses on domestic solutions, whose implementation – for example, in case of ex ante budget rules and changes in the budgetary process – and consequently its effectiveness still lies in the hands of elected politicians. What is missing in this literature is a fourth solution, namely the supervision of fiscal performance by supranational NMIs, such as the European Commission's role in the SGP, as the next chapter shall discuss in further detail (Baerg and Hallerberg 2016, 968–971; Köhler and König 2015, 334).

4 Supranational Non–Majoritarian Fiscal Surveillance in the European Union: Explaining (Non)–Compliance

Figures 3 and 4 [1]) show the development of deficit and debt levels of all EU countries for the time period from 2002 to 2017. The development is similar to that of OECD countries discussed before, as most EU countries maintain high deficit and debt levels. The pattern is particularly interesting since the time period shown in both figures is one in which the SGP came into force. Given the accumulation of high deficit and debt levels in many EU countries and the outbreak of the European sovereign debt crisis, the question about the effectiveness of supranational NMI fiscal surveillance in the EU has gained increased interest (Baerg and Hallerberg 2016; Köhler and König 2015; Köhler et al. 2018; Stéclebout-Orseau and Hallerberg 2007; van der Veer and Haverland 2018; Wijsman and Crombez 2017).

[1] Figures 3 and 4 have been created using the `R`–packages `car` and `ggplot2` (Fox and Weisberg 2011; Wickham 2016).

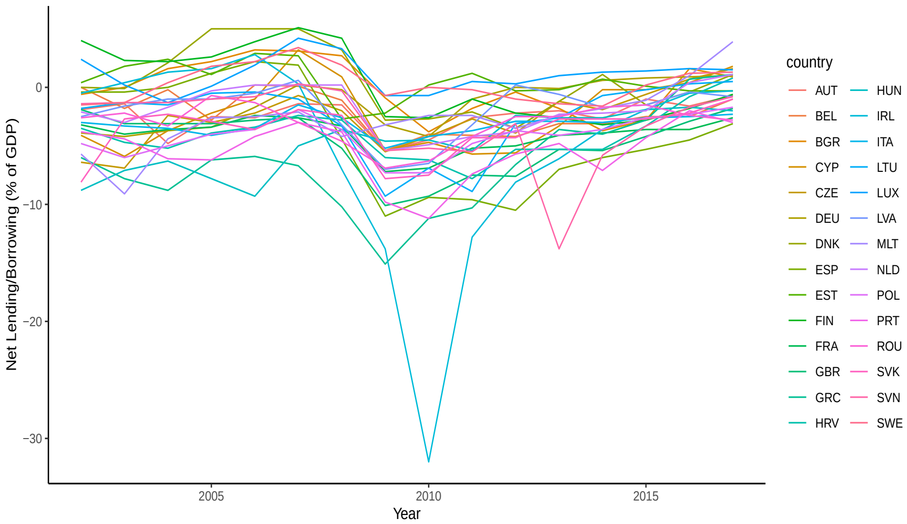


Figure 3: Budget Deficits of EU Countries since 2002

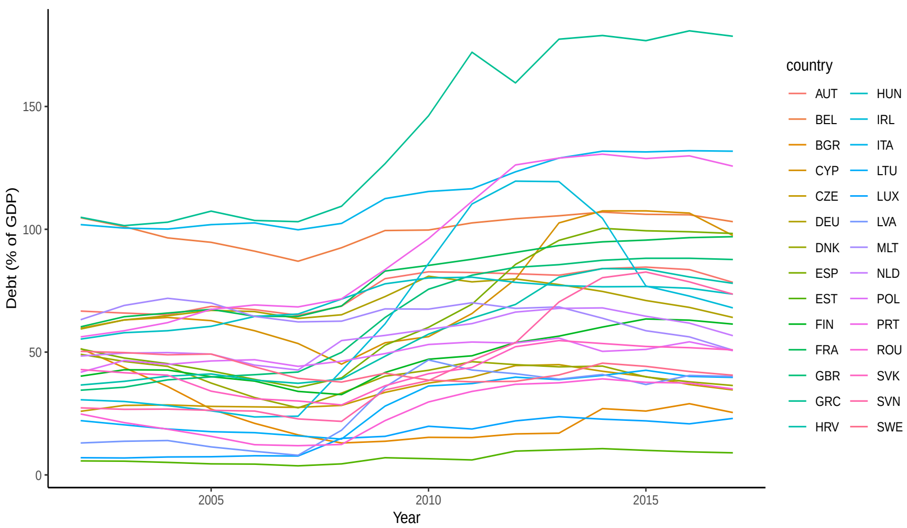


Figure 4: Debt Levels of EU Countries since 2002

Generally, the SGP and the European Commission as the SGP's "watchdog" were intended to serve as European–level institutions to maintain economic stability and convergence within the EU and particularly in the Eurozone (Baerg and Hallerberg 2016, 968–971; Köhler and König 2015, 334).

The purpose of the SGP is to avoid "spillover effects" that could threaten the common currency (Buti and Giudice 2002, 824; Faini 2006, 446–447). In a currency area such as the Eurozone, fiscal indiscipline of one member can affect the common interest rate and consequently the interest rates and fiscal policies of the other members (Buti and Giudice 2002, 824; Faini 2006, 446–447). Another problem is that given an expansionary fiscal policy or extensive public debt, the central bank is no longer able to guarantee price stability (Buti and Giudice 2002, 824). Furthermore, the SGP was established to avoid the independent European Central Bank facing pressures from undisciplined member states (Buti and Giudice 2002, 824–825). For these reasons, the SGP is an essential device to enhance the functioning of the EMU.

Article 126 and Protocol (No 12) on the EDP annexed to the Treaty on the Functioning of the European Union (TFEU) set out the main criteria regarding a) when a rule violation of the SGP takes place and b) what the Commission is required to do in case of such a violation (European Union 2008; Protocol (No 12) on the Excessive Deficit Procedure 2008).

According to Protocol No. 12 on the EDP, the reference values of the SGP are "3% for the ratio of the planned or actual government deficit to gross domestic product at market prices" and "60% for the ratio of government debt to gross domestic product at market prices" (Protocol (No 12) on the Excessive Deficit Procedure 2008). If a member state's deficit and/or debt level exceeds these values, the requirements of the pact are considered as not fulfilled (European Union 2008). In this case, the Commission is required to prepare a report considering whether the deficit exceeds government investment expenditure as well as the member state's medium–term economic and budgetary situation (European Union 2008). However, the pact allows for violating the reference values if the deficit ratio "[...] declined substantially and continuously and reached a level that comes close to the reference value," and the debt ratio "[...] is sufficiently diminishing and approaching the reference value at a satisfactory pace" (European Union 2008). Moreover, the Commission is asked to prepare a report if it fears the risk of an excessive deficit in a member state, even if the deficit and debt criteria are fulfilled (European Union 2008). This report on the Commission's opinion of the existence of an excessive deficit is the first step of the sanctioning process for SGP violations, the so–called "Excessive Deficit Procedure" (European Central Bank 2003, 59; Protocol (No 12) on the Excessive Deficit Procedure 2008).

The effectiveness of the SGP as well as the Commission's ability to enforce member states to comply with it have been investigated by several studies.

Generally, at the early stages of the EMU, the signing of the Maastricht treaty and thus the acceptance of its monetary and fiscal convergence criteria led to a reduction of deficits in high–deficit countries (Baskaran and Hessami 2017, 254–268). Interestingly, the deficits in these countries rose once again after the Euro was introduced (Baskaran and Hessami 2017, 268).

Regarding debt, the effect of the SGP also seems to vary across countries: while the SGP reduced public debt in Germany, the Netherlands, Belgium, France, Finland and Austria, Greece, Portugal and Italy were able to increase their public debt through membership in the monetary union (Köhler and König 2015, 1–27). This suggests that there exist free–riding and moral hazard problems within the Eurozone (Köhler and König 2015, 1). Even worse, von Hagen and Wolff (2006) find that the SGP induced "creative accounting" to lower deficits among memberstates, thereby leading the purpose of the pact to absurdity (von Hagen and Wolff 2006, 3259–3261).

The impact of the SGP also seems to depend on domestic characteristics: regarding the budgetary process, "commitment countries" in which the finance minister has a leading role in the budgetary decision–making process (such as Finland and the Netherlands) and "delegation countries", where different parties agree on a fiscal contract (e.g. France, Germany, the UK, Italy, Greece, Austria and Spain) show different behavior (Annett 2006, 5–6). "Commitment countries" performed better under the SGP than "delegation countries" (Annett 2006, 5–26). Furthermore, countries suffering from high growth volatility and small countries adopted more disciplined fiscal policies under the SGP (Annett 2006, 26).

Another problem is that powerful member states were able to weaken the pact (Baerg and Hallerberg 2016, 968; Heipertz and Verdun 2010, 149–150). During the first crisis of the SGP, France and Germany – which had not met its criteria in 2003 – strongly opposed the pact and its sanctions (Heipertz and Verdun 2010, 129–151). In addition, the indifference about an enforcement of the pact of the other two large member states, Italy and the UK, prevented the opening of EDPs against France and Germany, whereby consequently, the pact was weakened (Heipertz and Verdun 2010, 146–151). This weakening had an effect on public borrowing within the EU since both high– and low–deficit countries increased their deficits after the failure to sanction France and Germany (Baskaran and Hessami 2017, 268).

Powerful and Eurosceptic member states were also able to weaken the Commission's assessments of fiscal performance, which reduced member states' commitments to the EU (Baerg and Hallerberg 2016, 968–971). In addition, the strong bargaining position of the larger member states who had already violated the pact during its reform in 2005 and the lack of enforcement power of the Commission reduced the likelihood of sanctioning and therefore the effectiveness of the SGP (Eichengreen 2005, 438). However, van der Veer and Haverland (2018) find that the recent reforms aiming to prevent another "debt crisis" have strengthened the Commission's role as a "guardian of the markets" rather than the "people" (Bauer and Becker 2014, 219; van der Veer and Haverland 2018, 2–16). This is because higher politicization in member states leads to a) more country–specific recommendations (CSRs) addressing economic and fiscal risks in a given member state, and b) fewer recommendations for social investments (Bauer and Becker 2014, 219; van der Veer and Haverland 2018, 2–16).

Overall, the pact lacks powerful enforcement mechanisms in its institutional design since the biased Council of Finance Ministers has long been mainly responsible for enforcing the rules and large, influential countries are less likely to comply with the pact's fiscal policy rules (De Haan et al. 2004, 257). Although the introduction of the so–called "reversed qualified majority"[2] in the course of the *Six–Pack* reform increases the possibility for earlier sanctions, it is unclear whether it will solve the challenging task of enhancing the fiscal discipline of sovereign member states (Bauer and Becker 2014, 219–220; Buti and Carnot 2012, 906–908).

Finally, there is limited room for possible improvements of the pact. For example, the exclusion of "fiscal sinners" from the voting procedure would improve the effectiveness of punishing fiscal free–riders (Irlenbusch et al. 2003, 650–662). On the other hand, the introduction of a communication phase would reduce the prevalence of excessive deficits and ensure strong cooperation, but would have no effect on the frequency of sanctioning (Irlenbusch et al. 2003, 651–663).

In sum, both economists and political scientists find evidence that the SGP and its enforcement by the Commission were not successful in enhancing member states' compliance with the fiscal rules of the EU, and that even improving the pact is rather difficult. This evidence is supported by some descriptive statistics.

[2]The "reversed qualified majority rule" established that member states have to vote *against* sanctions and no longer *for* them if the SGP is violated (Bauer and Becker 2014, 219–220; Buti and Carnot 2012, 906–908).

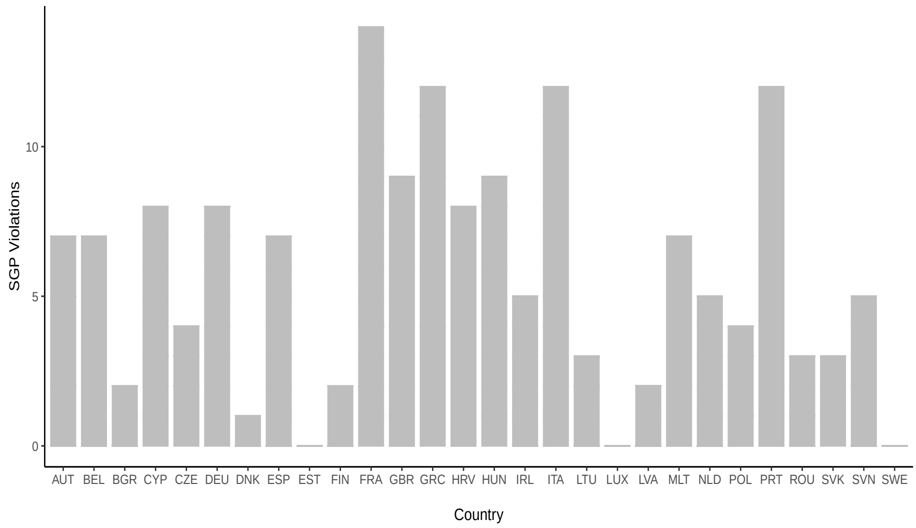


Figure 5: Number of SGP Violations per Country since 2002

Figure 5 [3] illustrates that all EU countries have broken the SGP rules (either the deficit or debt rule) several times, with Denmark having the lowest number of violations (1), while many other states – including large member states like Germany, France and Italy – committed up to sixteen violations[4]. However, these violations rarely result in sanctions: as Table 1 [5], shows, a violation of the SGP's rules (either > 3% of GDP or a debt level > 60% of GDP) only led to the opening of an EDP in 29 out of 144 cases [6]. Overall, these numbers confirm the impression of ineffective sanctioning by the Commission.

[3]Figure 5 has been created with `ggplot2` (Wickham 2016).

[4]The statistics presented in both the graph and table are based on a dummy variable that takes the value of one if either or both a country's deficit or debt level exceeds the pact's respective reference values of 3 or 60 percent and either ratio does not follow a diminishing trend, as defined in Article 126 of the Treaty on the the Functioning of the European Union (European Union 2008).

[5]Table 1 has been created using the `R`–package `descr` (Aquino et al. 2016).

[6]EDP has been measured as the occurrence of the Commission's opinion on an excessive deficit for a given country and year.

Table 1: Number of Sanctioned and Non–Sanctioned SGP Violations since 2002

	EDP		
SGP Violation	0	1	Total
0	253	9	262
1	115	29	144
Total	368	38	406

Thus, most studies on the SGP have been concerned with either a) its "top–down" effectiveness in terms of deficit or debt reductions and possible explanations for this or b) indirect mechanisms that affect the Commission's (sanctioning) behavior such as member states' influence on CSRs (Baerg and Hallerberg 2016, 968–971; Börzel 2003b, 1–2; van der Veer and Haverland 2018, 1). Regrettably, no studies investigated any direct "bottom–up" determinants of the Commission's (non–)sanctioning behavior via opening an EDP, the SGP's enforcement mechanism (Börzel 2003b, 1–2; European Central Bank 2003, 59; Protocol (No 12) on the Excessive Deficit Procedure 2008).

Even though it used to be the Council that has to agree on imposing sanctions on a litigant member state[7], it is the European Commission that – as the EU's central monitoring agency – is responsible for enforcing compliance with the SGP (Bauer and Becker 2014, 215–222; De Haan et al. 2004, 257; König and Mäder 2014; Köhler et al. 2018, 1–2). As written down in the Treaty on the Functioning of the European Union, the Commission takes the first initiative step towards these sanctions by initiating an EDP via delivering the "Commission's opinion on the existence of an excessive deficit" (European Union 2008; De Haan et al. 2004, 257; European Central Bank 2003, 59; European Commission 2017b; Köhler et al. 2018, 1–2; Protocol (No 12) on the Excessive Deficit Procedure 2008). However, given the pattern shown in Table 1, the question is: Why does breaking the SGP's rules not lead to consequent sanctioning by the Commission? The theoretical framework presented in the following part will provide an answer to this question.

[7]The so called *Six–Pack* reform introduced the principle of "reverse qualified majority voting", which means that from now onwards a qualified majority in the Council has to vote *against* sanctions rather than *for* sanctions as in the past, thereby strengthening the Commission's role in monitoring and sanctioning non–compliance with the SGP (Bauer and Becker 2014, 215–222).

5 Theoretical Framework – Determinants of Non–Majoritarian Decision–Making

The European Commission is responsible for fiscal surveillance and enforcing compliance with the SGP if a member state's budget exceeds the pact's deficit– or debt reference value. In this section, I explain that the Commission struggles with this task because it faces a legitimacy and a transparency problem at the member state level which stems from its design as a supranational NMI. For this purpose, I first briefly review two theoretical models on NMI decision–making and their most important implications. Based on these implications, I will derive my argument on what determines the Commission's ineffective sanctioning within the SGP.

5.1 The Legitimacy Problem of the European Commission

The European Commission is appointed by member state governments to act as the EU's executive in general and an enforcer of compliance with the rules of the SGP within EMU in particular (Baerg and Hallerberg 2016, 968–972; Follesdal and Hix 2006, 535–536; Köhler et al. 2018, 9). As such, it lacks any electoral legitimacy since it is neither directly or indirectly elected (Follesdal and Hix 2006, 535–536). The key problem in the Commission's institutional design is that its legitimacy is limited compared with member state governments it is supposed to sanction. This is problematic in the context of the SGP since the Commission's lack of legitimacy is assiciated with a lack of authority to enforce the compliance of elected governments.

Because the Commission is not directly elected but equipped with public authority, it can be classified as a NMI (Thatcher and Sweet 2002, 1–2). These institutions are known to rely on public support to gain legitimacy and consequently the necessary authority to sanction governments (Caldeira and Gibson 1992, 695; Köhler et al. 2018, 11–15; Vanberg 2005, 19–24).

In order to illustrate the role of public support for the Commission's sanctioning behavior, Köhler et al. (2018) set up a bargaining game between the European Commission and EU member state governments about compliance with the SGP's fiscal rules. The model provides predictions about when the Commission will launch an EDP to sanction member states that have broken the rules of the SGP depending on voters' support for compliance with the SGP.

The model is set up as follows: the main actors are the European Commission and a member state government. In the first stage of the game, the government decides on a budget, whose compliance with the SGP's rules is evaluated by the Commission (Köhler et al. 2018, 16–18). Naturally, if the budget is in line with the SGP, the game ends (Köhler et al. 2018, 20–21). If not, the Commission has to decide whether to open an EDP or not, and if it does not, the game ends (Köhler et al. 2018, 20–21). After the initiation of an EDP and the public's reaction to it, the government makes the decision about whether to adjust the budget or not (Köhler et al. 2018, 21). If it adjusts B, the game ends (Köhler et al. 2018, 21). Otherwise, the Commission has to decide about escalating the conflict or not, where the government will only endure costs in the former case (Köhler et al. 2018, 21).

Note that the decision to initiate an EDP at the second stage of the game is not a simple one based only on whether the reference values of the pact are violated; rather, the Commission has to take into account that – in contrast to itself – the government is directly elected and represents the interests of its voters.

The crucial aspect of the model for understanding the sanctioning behavior of the Commission is that the bargaining position of both the government and the Commission is determined by voters' relative preferences in terms of overspending due to consumptive interests versus normative considerations about fiscal responsibility, namely compliance with the SGP (Köhler et al. 2018, 4–22).

The government is affected by voters' preferences given that it will suffer from "audience costs" if it does not support voters' interest (Köhler et al. 2018, 17). If voters are more in favor of fiscal responsibility than consumption based on (over)spending, the government faces high audience costs due to a credibility loss caused by the violation of the SGP's rules, which can remove it from office (Köhler et al. 2018, 17–18). This threat will then make it comply with the rules of the SGP, whereby the government is in a "weak" bargaining position (Köhler et al. 2018, 17–18). By contrast, if most voters prefer overspending against fiscal consolidation, the government's audience costs and thus the threat of being removed from office are low even if it does not follow the rules of the SGP (Köhler et al. 2018, 17–18). In this case, the government is in a "strong" bargaining position (Köhler et al. 2018, 17–18).

Similarly to the government's audience costs, the Commission will face "enforcement costs" that stem from a loss in credibility if it fails to achieve compliance (Köhler et al. 2018, 18). It faces low enforcement costs if the government is in a "weak" bargaining position since voters prefer fiscal responsibility over non–compliance (Köhler et al. 2018, 18). However, if the government has a "strong" position in the compliance conflict because most voters prefer overspending against fiscal consolidation, the Commission is confronted with high costs when enforcing compliance with the SGP (Köhler et al. 2018, 18). In sum, the Commission's enforcement costs are inversely related to the type of the government, where a weak government bears low and a strong government high enforcement costs (Köhler et al. 2018, 4–19).

Besides enforcement costs, the Commission also faces "reputation costs" that are imposed by other governments if it fails to enforce compliance with the SGP's fiscal rules (Köhler et al. 2018, 18). This is because failed enforcement can motivate other governments to overspend and also violate the SGP's rules (Köhler et al. 2018, 18–19). The amount of reputation costs is related to the probability of succeeding in a compliance conflict which depends on the government's type (Köhler et al. 2018, 18–22). Naturally, the Commission will only escalate the conflict in the last stage of the game only against "weak governments" (where public support for fiscal consolidation is high) since it can always expect the government to comply in this case (Köhler et al. 2018, 22–23). This implies that if public support for fiscal responsibility is high, both the Commission's enforcement costs and reputation costs are low. In equilibrium, the Commission will only initiate an EDP if the gains from compliance are larger than both the enforcement and reputation costs imposed from other countries; otherwise, it will refrain from doing so (Köhler et al. 2018, 24–25). Recall that the Commission's enforcement costs are high if it faces a "strong" government where the majority of voters prefers overspending over fiscal responsibility (Köhler et al. 2018, 18). Moreover, its reputation costs depend on the probability of success which is higher in case of a “weak” government where public support for fiscal consolidation is high (Köhler et al. 2018, 22–23). Overall, the predictions of the model imply that the initiation of an EDP by the European Commission is mainly driven by the relative amount of voters’ support for compliance with the SGP versus support for overspending.

5.2 The Transparency Problem of the European Commission

The model developed by Köhler et al. (2018) stresses the importance of public support for the Commission’s decision to open an EDP. However, the Commission faces another difficulty when confonted with member state non–compliance. Due to its supranational nature, it has imperfect information about whether an excessive fiscal policy on the member state level is justified – for example, in the occurrence of a sudden economic crisis – or is exclusively caused by serving voters’ consumptive interests (Köhler et al. 2018, 5–15). Moreover, this imperfect information is made worse in the case of creative accounting (von Hagen and Wolff 2006).

In that sense, the government faces a trade–off between responsibility and responsiveness (Köhler et al. 2018, 2–7). Responsibility broadly refers to making decisions while considering their consequences (Pennock 1952, 796–797). In the context of the SGP, it means (long–term) fiscal responsibility by complying with the rules of the pact (Köhler et al. 2018, 2–7). On the other hand, responsiveness can be defined as “responding easily to any and all demands” (Pennock 1952, 791). In terms of (non)–compliance with fiscal rules, it refers to meeting the (short–term) consumptive demands of voters (Köhler et al. 2018, 2). Consequently, a responsible government will comply with the SGP (Köhler et al. 2018, 2–9). However, in case of non–compliance and the decision to initiate an EDP, the Commission has to decide whether the government is merely responsive to voters’ (consumptive) demands or whether it spends more due to economic trouble, which would be the case in the presence of an economic shock (Köhler et al. 2018, 11–15). Since only the government has perfect information on the true nature of non–compliance, the Commission is uncertain when deciding about sanctioning overspending member state governments (Köhler et al. 2018, 5–15). If it sanctions a state that already faces a recession, the Commission would make itself even more unpopular among voters and further undermine its legitimacy (Köhler et al. 2018, 5–15). Besides a lack of legitimacy, the Commission therefore also faces lack of transparency.

In order to understand the role of transparency for the Commission's sanctioning behavior in the context of the SGP, it is helpful to consider theories on NMI decision–making from judicial politics. Here, researchers early recognized the tension between a constitutional court, which is responsible to ensure that laws made by the government are in line with the country's constitution, and the government (Caldeira and Gibson 1992, 695; Vanberg 2005, 19–24). Similar to the Commission, the core problem of the court is that it is not legitimized directly via elections, which defines it as an NMI (Caldeira and Gibson 1992, 695; Thatcher and Sweet 2002, 2); Vanberg 2005, 19–24).

Vanberg (2005) develops a simple game–theoretic model to explain the interaction between a constitutional court and a government. His key argument is that the amount of public support directly translates into political power, which helps to enforce (judicial) decisions (Vanberg 2005, 19–20). The simple mechanism behind this is that ignoring a court's decision in the presence of citizens that regard judicial independence as important will lead to a loss of public support for the government and hence a loss of political power (Vanberg 2005, 19–20). This threat gives governments an incentive to comply with the court's rulings (Vanberg 2005, 19–20). However, even in case of non–compliance, governments will usually refrain from publicly admitting that they are not willing to implement the court's decision since they anticipate that most citizens value judicial independence (Vanberg 2005, 21–22). Rather, the government will try to simply ignore the court's ruling (Vanberg 2005, 21–22). In such a situation, it becomes clear that it is not sufficient that the public supports the court to make the government comply, it is also crucial that citizens recognize government evasion (Vanberg 2005, 21). Thus, recognition involves monitoring of the government's reactions to the court's decision and the activation of public support for the court in case of non–compliance of the government (Vanberg 2005, 21–22). This monitoring capacity, in turn, depends on the amount of transparency of the political environment (Vanberg 2005, 22–23). Note that not only the government bears some political costs, also the court does in case of governmental non–compliance since it undermines the court's authority (Vanberg 2005, 27). Therefore, the court is interested in the implementation of its decisions (Vanberg 2005, 27). Moreover, both the court and the government are uncertain about the levels of public support and transparency in their environment, so their behavior depends on these actors' subjective beliefs about the joint probability of these two conditions (Vanberg 2005, 24–28). In addition, the model considers the court's preferences, which can be either in line with the government ("convergent") or against them ("non–convergent") and will affect the behavior of the court as well (Vanberg 2005, 24–28).

From the equilibria that Vanberg (2005) derives from his model, those that predict the behavior of the non–convergent court conditional on the levels of the joint probability of public support and transparency hold particular interest (Vanberg 2005, 28–36).

If public support and transparency are low, a non–convergent court will not veto legislation since it anticipates evasion (Vanberg 2005, 32). If public support and transparency are high, the court will veto the law the government has passed, and the government will comply afterwards (Vanberg 2005, 34–35). Finally, if the joint probability of public support and transparency are at an intermediate level, a non–convergent court will veto the government's legislation despite expecting government evasion (Vanberg 2005, 35–37). In sum, Vanberg's (2005) model considers not only public support, but also transparency as a decisive influence factor on NMI decision–making.

The link of this model to the case of the European Commission is that both the court and the Commission can be classified as NMIs since they possess some separate public authority, but are not legitimized via elections. Moreover, similar to a court, the Commission can "veto" a governmental decision, albeit not on a law which might not be in line with the constitution, but on a budget that violates the rules of the SGP. Finally, the Commission can be seen as a permanent non–convergent NMI because as the "watchdog" of the SGP, it is always in favor of compliance and would therefore never prefer governmental overspending. As such, the Commission's sanctioning behavior can be expected to be similar to the one of a non–convergent constitutional court that is affected by the levels of both public support and transparency.

However, the transparency problem of the Commission is somewhat different from the one of the court: in Vanberg's (2005) model, the constitutional court is in need of a transparent *political environment* that provides citizens with information on government evasion, which then mobilizes public support and helps the court to finally achieve the government's compliance. In the case of the SGP, in contrast, I argue that the Commission is in need of transparency about the country's *economic situation* before opening an EDP to reduce its uncertainty about whether the government's SGP violation is due to voter responsiveness or not. Only if sufficient transparency about the respective country's economic conditions are given does the Commission not risk a loss in public support in the future caused by unjustified sanctioning.

The only possibility for the Commission to infer about the government's responsiveness to voters is to evaluate whether an excessive deficit or debt level is due to extraordinary circumstances such as an economic crisis. Article 126 of the Treaty on the Functioning of the European Union asks the European Commission to consider the economic situation of a country so that it "[...] shall examine compliance with budgetary discipline on the basis of the following two criteria: (a) whether the ratio of the planned or actual government deficit to gross domestic product exceeds a reference value, unless: [...] the excess over the reference value is only exceptional and temporary and the ratio remains close to the reference value;" and b) "whether the ratio of government debt to gross domestic product exceeds a reference value, unless the ratio is sufficiently diminishing and approaching the reference value at a satisfactory pace" (Buti et al. 1998, 84; European Union 2008). The pact further defines a sudden decline in real GDP by between 0.75 and 2 percent in the respective year as "exceptional and temporary", which allows for flexibility during recessions (Buti et al. 1998, 84; Eichengreen and Wyplosz 1998, 70).

Based on the legal framework of the pact, the Commission can conclude that overspending is *not* caused by unusual economic turbulences if the above–mentioned criteria of exceptionality and temporality are *not* met. In this case, transparency on the country's economic situation is given insofar that the Commission will use the absence of difficult economic circumstances as a signal that a SGP violation is due to the government's needs to serve voters' consumptive interests. Consequently, it will initiate an EDP without risking a further loss in public support.

5.3 Public Support and Transparency as Determinants of EDP Initiations

I have already stated above that the European Commission – like constitutional courts – can be defined as a NMI. I further argue that the Commission is confronted with a legitimacy and transparency problem which is caused by its design as a supranational NMI. The identification of these two problems allow me to draw the following arguments about the determinants of EDP openings by the Commission.

First, the level of voters' support for the Commission and/or the rules that it enforces – in either absolute or relative terms compared with the amount of support for the government – matters for its sanctioning behavior. Only if the Commission enjoys a sufficient amount of public support does it have sufficient legitimacy and therefore political power to sanction rule–violations by (elected) governments.

Second, public support is not sufficient to explain the sanctioning behavior of the Commission. Since the Commission is uncertain about the true nature of governmental non-compliance – namely whether it is caused by voter responsiveness or economic shocks – and would risk a future loss in public support if it sanctions a country that suffers from a recession regardless – it requires sufficient transparency about the economic situation of a country before opening an EDP. Table 2 provides an overview of the determinants of EDP initiations per theory.

Table 2: Overview: Determinants of EDP Initiations

Vanberg (2005)	**Köhler et al. (2018)**
Absolute Public Support for NMI	Relative Public Support for NMI
Transparency	

In sum, from the conclusions made above, I expect the Commission to open an EDP only if public support for compliance with the SGP is high and the respective country is not suffering from a recession. From this expectation, the following, testable hypothesis can be derived:

> H1: If there is public support for compliance with the SGP and transparency is high, the Commission will initiate an EDP to sanction SGP violations.

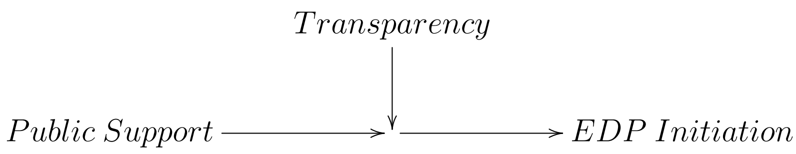
Having outlined the theoretical framework and the causal mechanism I am interested in, I will continue with the description of the research design to test Hypothesis 1.

6 Research Design

6.1 Potential Outcome Framework

In order to determine which quantities shall be estimated in this investigation, it is crucial to set up a so–called "potential outcome framework" as proposed by Rubin (2005) (Rubin 2005, 323–326). Based on this framework, the causal effect to be investigated here can be described as follows: let units (i) be countries that have violated the SGP, treatment (X) the amount of public support for the Commission for a country in a given year, outcome (Y) sanctions initiated by the Commission for that country and that given year (where outcomes in treatment states are Y_0 (= no sanctions), and Y_1 (= sanctions), Q an effect modifier indicating transparency about the country's economic situation, and Δ the treatment effect (Rubin 2005, 323–326; Pearl 2009, 133–136; VanderWeele and Robins 2007, 3). Each unit i (country) is only observed in one treatment state, i.e. $X_i = 0$ (low public support for the Commission) or $X_i = 1$ (absence of public support for the Commission) (Rubin 2005, 323). Moreover, each unit has one observed outcome, i.e. either Y_{i0} (no sanctions) or Y_{i1} (sanctions). Furthermore, one only ever observes $Y_i = X_i Y_{i1} + (1 ˘ X_i) Y_{i0}$ (Rubin 2005, 323). The causal effect of the treatment X (*Public Support*) on outcome Y (*EDP initiation*) comparing two levels of the treatment ($X = 0$ and $X = 1$) given a modifier Q (*Transparency*) is defined as $Y_{i(1)}|Q = q ˘ Y_{i(0)}|Q = q$, where the causal effect of X is not constant in q (Rubin 2005, 323; VanderWeele and Robins 2007, 3). Both the treatment and the moderation effect as described by Hypothesis 1 are illustrated in Figure 6 .

Figure 6: Causal Effect to be Investigated According to Hypothesis 1



Generally, I am interested in the average treatment effect of public support on sanctions initiated by the Commission given the absence/presence of transparency about the country's economic situation. However, for the analysis it is crucial to determine whether the third variable transparency acts like a *moderator* or a *mediator* since each requires a different modeling technique (Baron and Kenny 1986, 1174–1177). According to Baron and Kenny (1986), a moderator is *not* strongly related to both the treatment and the outcome, while Kraemer et al. (2002) propose to test for correlation between treatment and moderator to distinguish it from mediators (Baron and Kenny 1986, 1177–1178; Kraemer et al. 2002, 879–881).

In order to determine whether transparency acts a as a moderator or a mediator, I assessed the relation between each transparency indicator and the treatment variable public support as well as the two measures of the outcome variable EDP initiation, respectively. Unfortunately, since the moderator/mediator variable transparency (measured by real GDP and GDP growth) is continuous, while the dependent variable EDP initiation is a binary variable, a correlation analysis is not possible due to the different scaling of these variables (Gehring and Weins 2009, 141). Instead, I checked the value of η^2, a measure of strength of association between a nominal and a continuous variable [1], between the binary EDP indicators and real GDP (Gehring and Weins 2009, 141; Pierce et al. 2004, 917). η^2 ranges from 0 to 1 and allows me to evaluate whether transparency acts as a moderator or a mediator since higher values indicate a stronger association between two variables (Gehring and Weins 2009, 141; Pierce et al. 2004, 917–918). The association between the transparency and the outcome variable EDP proved to be very weak (close to 0).

Because both the treatment variable public support and the mediator/moderator variable transparency are measured by continuous variables, I conducted a simple Pearson's correlation analysis between each public support– and transparency indicator, respectively (Gehring and Weins 2009, 165–170). Here, Pearson's r only indicated a weak to medium

[1]The value of η^2 was assessed by using the R–package `BaylorEdPsych` (Beaujean 2012).

correlation (Gehring and Weins 2009, 170). Due to the absence of a strong correlation of transparency to both the treatment and the outcome, transparency can be defined as a moderator, not as a mediator (Baron and Kenny 1986, 1177–1178).

6.2 Data

The data sources used for this analysis are the Quality of Government dataset (version of January 2019), the Eurobarometer survey, Eurostat, the International Monetary Fund (IMF), the World Bank and the website of the European Commission and the Organization for Security and Co–operation in Europe . The data from these various sources are combined in a net dataset[2]. I will focus on all EU countries since the SGP applies to all EU states, not only to the Eurozone members (European Commission 2017d). Countries that have not been part of the EU at a given year yet are excluded from the analysis because the SGP is not applicable to them in this case.

As the time period to be investigated, I choose the period from 2002 to 2017 since this is the longest period for which data on the dependent and independent variables are available since the introduction of the Euro in 1999 (European Commission 2017b; European Union 2018b). Moreover, because I aim to explain variation in EDP initiations by the Commission among "fiscal sinners" by public support and transparency, the dataset only comprises observations where a) a member country's deficit level exceeded 3% of GDP *and/or* its debt level was higher than 60% of GDP in a given year[3] and b) there exists an *increasing* trend of the deficit or debt ratio at the same time (European Union 2008) [4].

[2]The following `R`–packages have been used for data preparation: `car`, `dplyr`, `countrycode`, `foreign`, `gdata`, `readstata13`, `readxl`, and `lubridate` (Arel-Bundock 2017; Fox and Weisberg 2011; Garbuszus and Jeworutzki 2016; Grolemund and Wickham 2011; R Core Team 2015; Warnes et al. 2017; Wickham and Bryan 2017; Wickham et al. 2017).

[3]In line with Protocol No. 12 annexed to the Treaty on the Functioning of the European Union (TFEU), deficit is measured as annual net lending/borrowing (current and capital account) as a percentage of GDP, while debt is measured as general government gross debt as a percentage of GDP (Protocol (No 12) on the Excessive Deficit Procedure 2008; Eurostat 2018); International Monetary Fund 2019b).

[4]The diminishing trend is coded based on two variables, $\Delta_{Deficit}$ and Δ_{Debt}, that measure the difference of the Deficit/Debt ratio between year t and year $t\breve{}1$ (Kohler and Kreuter

The consideration of the two trend criteria besides the reference values for deficit and debt is necessary to account for the fact that the pact allows for no sanctions if a country's deficit and/or debt ratio has declined before, even if they exceed the pact's reference values, and to account for countries that remain shortly below the reference values despite rising deficit and debt levels (European Union 2008).

The dependent variable is a binary variable taking the value of 1 if an EDP was opened against a given country in a given year, and 0 otherwise. I refrain from coding years with an ongoing EDP as 1 since I focus on the determinants of the *initiation* of an EDP, which is up to the Commission, and not their continuation, which is decided by the Council (European Central Bank 2003, 59). For the countries and the time period to be investigated, 29 EDPs occurred[5].

Note that studies on Commission decision–making are usually prone to self–selection bias as the Commission only publishes cases of successful sanctioning (Fjelstul and Carrubba 2018, 429–437). However, this investigation is less likely to suffer from such a self–selection bias as no EDP initiation by the Commission so far led to the imposition of fines by the Council, which would be the analogy to a court ruling against a litigant member state (European Commission 2017b).

Hypothesis 1 postulates an impact of public support for the Commission on EDP initiations conditional on the moderator transparency. Because moderators can be included in the analysis via interacting it with the treatment, the key independent variables of interest are six interaction terms consisting of each public support– and transparency indi-

2008, 79). A negative value of $\Delta_{Deficit}$ indicates that a country's deficit increased in year t compared with the previous year. By contrast, a positive value of Δ_{Debt} means that the debt ratio to GDP of a country increased at time t. However, note that the terms "closeness to the reference value" and "sufficiently diminishing and approaching the reference value at a satisfactory pace" as referred to in Article 126 for both a country's deficit and debt level are not defined. Since this lack of definition makes these two conditions unmeasurable, they can not be considered when creating the dataset on cases of SGP violations (European Union 2008; (Eichengreen and Wyplosz 1998, 70).

[5]According to King and Zeng (2001) an event can be considered as rare event if less than 5% of the data indicate its occurrence (King and Zeng 2001, 157). For a N of 144 in the dataset used for the statistical analysis (SGP violations only), a number of 7.2 EDPs would be equal to 5% of the data. However, since the occurrence of 29 EDPs exceeds 5% of the observations in each dataset, I will not apply any rare events specification.

cator,respectively (Baron and Kenny 1986, 1174; Brambor et al. 2006, 64; Kraemer et al. 2002, 880). Following Ai and Norton (2003), there will be a separate model for each of the six interaction terms (Ai and Norton 2003, 126). Besides the interaction terms, its constituent terms will be included in the model to avoid an omitted variable bias of the estimates (Brambor et al. 2006, 66–67).

In addition, I will control for possible confounding variables (Pearl 2009, 113–114). These are influential member states, non–compliance of other member states, the number of infringement procedures in a given year per country, Eurozone membership, SGP reforms implemented in the aftermath of the EU's financial and economic crisis and elections. The operationalization of these variables as well as the reason for their inclusion will be presented in the following.

6.2.1 Variables and Operationalization

Dependent Variable and Independent Variables

Dependent Variable – Commission's Initiation of an EDP: For the EDP procedure, two steps can be distinguished (European Union 2008; European Central Bank 2003, 59). At the beginning, there is the assessment stage, where the EDP is initiated (European Union 2008; European Central Bank 2003, 59). This stage involves the writing of a report by the Commission if a given country violates the Pact's reference values or is at risk of an excessive deficit (European Union 2008; European Central Bank 2003, 59). The Commission further decides on whether it will send an opinion to the Council on the existence or risk of an excessive deficit (European Union 2008; European Central Bank 2003, 59). If it does, the Council finally decides whether the country's deficit is indeed excessive (European Union 2008; European Central Bank 2003, 59). If it does not, the EDP is closed. In the case that the Council decides that a country's deficit violates the rules of the SGP, it formulates a recommendation for measures to correct the excessive deficit and sets a deadline when this correction has to be completed (European Union 2008; European Central Bank 2003, 59).

The second stage of the EDP procedure allows the Council to impose sanctions on the accusing memberstate (European Union 2008; European Central Bank 2003, 59). If the Council shares the opinion that the country has implemented the recommended measures, the EDP is suspended until the Council decides that the excessive deficit has been corrected in a timely manner (European Union 2008; European Central Bank 2003, 59). More procedural steps such as further recommendations, deadlines or new targets might follow if the country has not implemented the recommended measures or if they are insufficient (European Union 2008; European Central Bank 2003, 59; European Commission 2017a)[6]. Finally, in the case of overall non–compliance, the Council can sanction the reluctant country by a deposit and eventually a fine (European Union 2008; European Central Bank 2003, 59).

From the EDP's procedural stages we can see that the European Commission's role is reduced to the assessment step of the procedure, where it takes two decisions: the first one is to identify non–compliance and write a report and the second one on the existence of an excessive deficit that is addressed to the Council, which leads to the continuation of the EDP [7]. Due these two decision–making steps, two codings for EDP initiations will be used: first, a dummy variable that takes the value of 1 if the Commission publishes a report about a given country in a given year that breached or is at risk to violate the rules of the SGP ("EDP–Report"); and second, a dummy variable taking the value of 1 if the Commission's opinion on the existence of an excessive deficit is present for a given country and year, and 0 otherwise ("EDP–Opinion"). Because only the transfer of such an opinion to the Council marks the Commission's final decision on the existence of an excessive deficit in a given country, the primary analysis will be run using the "EDP–Opinion" variable, while the "EDP–Report"–specification will be used to assess the robustness of the results (European Union 2008). Both codings are based on the overview of EDPs published on the Commission's website (European Commission 2017b).

Key Independent Variable 1 – Public Support: Public support for the European

[6]The scope and frequency of monitoring the effectiveness of the measure implementation has been increased by the Two–Pack regulation (European Commission 2018a). Now, the concerning member state has to report its progress every three to six month in order to ensure the early detection of implementation failure before the deadline set by the Council (European Commission 2018a).

[7]Note that in contrast to the opening of infringement procedures, such as investigated by Fjelstul and Carruba (2018), the Commission will *not* announce an opinion on the existence of an excessive deficit to the Council if the respective country has not taken any action to reduce its deficit and/or debt level after the publishment of the report (Fjelstul and Carrubba 2018, 431–437). Therefore, a conditional coding as a means to overcome self–selection problems such as done by Fjelstul and Carruba (2018) is not possible.

Commission/compliance with the SGP will be operationalized by three indicators: first, the *absolute* level of trust in the European Commission based on the respective Eurobarometer item (Vanberg's (2005) interpretation, see Table 2), second, the *relative* level of trust in the Commission relative to trust in the national government (Köhler et al.'s (2018) interpretation, see Table 2), that are both based on the respective Eurobarometer items, and third, the Eurobarometer item asking whether the respondents have a positive image of the EU on a scale (European Commission 2018b, European Commission 2018c). I explain these indicators in further detail in the following paragraphs.

Public support for the Commission's policy is operationalized as trust in the Commission here because people are more likely to accept decisions and policies – such as the rules of the SGP – by institutions that they trust as they consider them to be legitimate (Marien and Hooghe 2011, 268). Since the theory of Vanberg (2005) focuses on the *absolute* level of public support for the NMI only , the absolute level of trust in the Commission is used as the first indicator in this investigation (Vanberg 2005, 19–49). In order to determine the degree of trust in the Commission in each country, the variable measuring the level of trust in the Commission taken from the Eurobarometer survey will be recoded as a dummy variable, where 1 indicates trust and 0 no trust. "Don't know" answers are coded as an extra category. In order to determine the degree of trust in the Commission in each country, I will then compute the share of individuals that expressed trust in the Commission per country.

In contrast to Vanberg (2005), Köhler et al. (2018) highlight the importance of the *relative* level of public support as the Commission's decision to initiate an EDP is based on whether most voters support compliance with the SGP or overspending (Köhler et al. 2018, 17–25). Therefore, the level of public support for the national government's overspending policy besides the level of public support for the Commission needs to be taken into account. Public support for governmental overspending will be measured by the Eurobarometer item asking for the degree of trust towards the national government. I choose this indicator because high trust in the national government increases public support for government (over)spending (European Commission 2017c); Rudolph and Evans 2005, 669). As for public support for the European Commission, the share of individuals that expressed trust in the national government per country is computed.

In order to measure the relative amount of public support for compliance with the SGP compared with the amount of support for overspending, I generate a variable where the share of individuals that trust the national government is subtracted from the share of individuals expressing trust in the Commission for each country and year (Kohler and Kreuter 2008, 79). If public support for the Commission (= support for compliance with the SGP) is relatively higher than support for the national government (= support for overspending), then the variable takes a positive value. The higher the Commission's support vis–à–vis support for the national government, the higher is that value and consequently, support for compliance with the SGP's fiscal rules. In that case, the probability for the initiation of an EDP is expected to increase if a country's deficit and/or debt level is not in line with the SGP rules.

Finally, the possibility that people mistrust the Commission since it is often perceived as a "bureaucracy", but generally support the EU and its regulations, needs to be taken into account (Christiansen 1997, 76). For this purpose, I follow McLaren (2007) and De Vries and van Kersbergen (2007) and use the Eurobarometer item asking for a positive image of the EU to check the robustness of the results (McLaren 2007, 241; De Vries and Kersbergen 2007, 314). Since the indicator measuring whether the interviewed person has a positive image of the EU is a scale variable ranging from 1 (= "very negative") over 3 (= "neutral") to (4) (= "fairly positive") and 5 (= "very positive"), I compute the share of individuals that stated to have a "very" or "fairly" positive image of the EU per country from this variable (European Commission 2018b). As for the previous public support indicators, this variable will be interacted with each transparency indicator.

Key Independent Variable 2 – Transparency: As explained in the theory section above, transparency refers to the situation where the Commission can infer that a violation of the SGP's deficit and/or debt rule is due to voter's consumption interests and not caused by an unexpected economic shock (Köhler et al. 2018, 5–15; Vanberg 2005, 45–49). The pact sees an "exceptional and temporary" deficit as in case of a recession as given if there is a sudden decline in real GDP in the respective year (European Union 2008; Buti et al. 1998, 84; Eichengreen and Wyplosz 1998, 70). If these criteria for exceptionality and temporality are not met, the Commission will know that the government violates the pact to serve voters' consumption preferences and initiate sanctions. As mentioned in the pact, transparency will be measured by real GDP (constant 2010 U.S. dollar) (The World Bank Group 2019). If a country violates the pact while having a rise in real GDP, it is obvious that a recession is absent and the probability of an EDP initiation increases.

Besides the decline in real GDP as defined by the pact, negative GDP growth is another indicator that a country suffers from an economic recession (Blanchard 1993, 2070). *Positive* growth rates, in turn, would consequently indicate that a government is violating the pact due to deliberate overspending. Therefore, also this variable will be used as a second transparency measure for robustness checks. GDP growth is measured by real GDP growth in percent (International Monetary Fund 2019c).

Covariates

Pearl (2009) proposes to select the covariates of an analysis according to the so called ‘back–door’ criterion (Pearl 2009, 113). This can be explained as follows. When investigating the causal effect of X (public support) on Y (sanctions/no sanctions), there are usually many additional measurable and unmeasurable factors around that influence either Y or X or both X and Y (Pearl 2009, 113). To account for this problem, it is necessary to select a subset of factors which values are held constant for both treated and untreated subjects (Pearl 2009, 113). This is called an ‘admissible’ or ‘sufficient’ set (Pearl 2009, 113–114). An ‘admissible set’ S can be defined according to two criteria: first, no element of S descends from treatment X, and second, the elements of S are blocking all “back–door” paths from X to Y, namely all paths ending with an arrow pointing to treatment X (Pearl 2009, 114). By conditioning for S, all “back–door” paths are blocked and it can be ensured that only the causal effect from *public support* to *EDP initiation* conditional on transparency, and no other spurious associations, will be measured (Pearl 2009, 114).

For this investigation, I identified the factors influential member states, non–compliance of other member states, the number of infringement procedures in a given year per country, Eurozone membership[8], SGP reforms implemented in the aftermath of the EU’s financial and economic crisis and national elections to be part of the admissible set S because both arrow emitting nodes of these two variables are in S (Pearl 2009, 114). Their influence on the causal effect of *public support* on *EDP initiation* conditional on high/low levels of *transparency* is illustrated in Figure 7 to 10 and shall be explained in further detail in the remainder of this section.

[8]Due to the application of a multi–level model in this investigation, it is noteworthy that the inclusion of time–constant factors such as Eurozone membership improves the estimation of the variance of the country–level units from the data (Shor et al. 2007, 176–177).

Figure 7: Effect of Powerful Member States

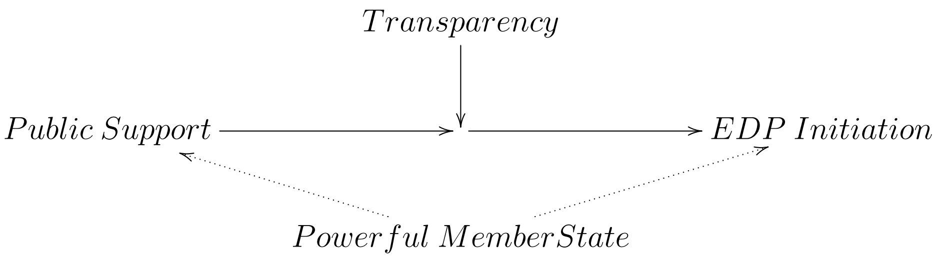


Figure 8: Effect of Non–Compliance of Other Member States

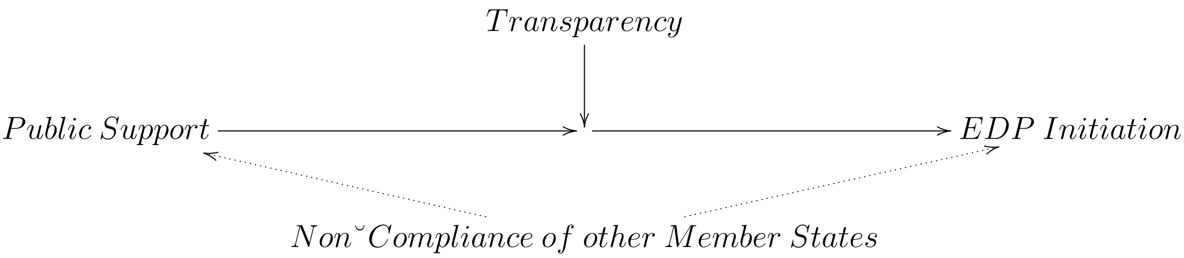


Figure 9: Effect of Eurozone Membership

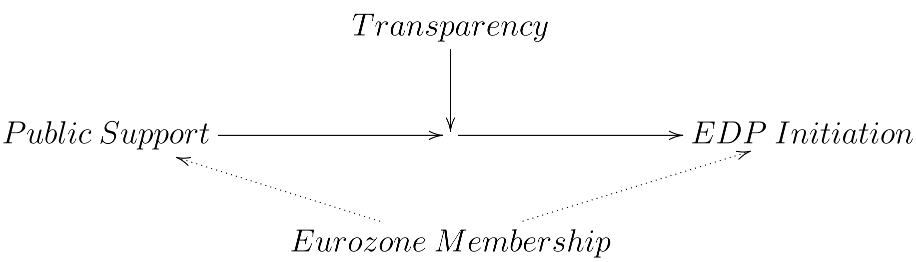
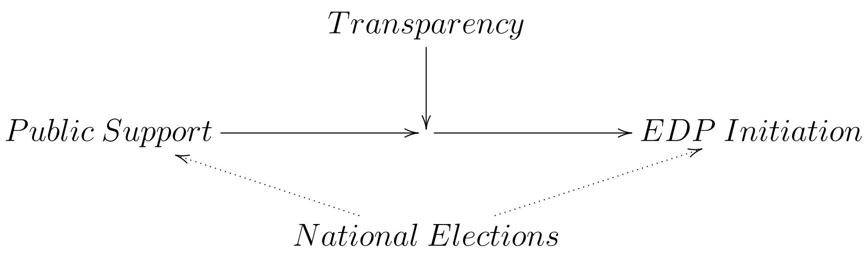


Figure 10: Effect of National Elections



Control Variable 1 – Powerful Member States: Influential member states – particularly those having a large economy – affect both public support and the probability of initiating an EDP by the Commission. One the one hand, voters in large member states have become more Eurosceptic since the strengthening of the European Parliament in the legislative process of the EU, the removal of an additional Commissioner from the larger member states and the enlargement of the EU by small countries have diminished their influence in the EU decision–making process (Hix 2007, 139–140). All of these changes have made it less likely that the large member states would always achieve their preferred policy outcome, whereby support for the EU consequently has decreased (Hix 2007, 140). This reduced support for the EU as a whole is likely to reduce support for compliance with the rules of the SGP.

On the other hand, Baerg and Hallerberg (2016) provide empirical evidence that powerful memberstates weakened the Commission's assessments of fiscal performance and consequently the pact's sanctioning mechanisms (Baerg and Hallerberg 2016, 968–972). Therefore, it is less likely that an EDP is opened against large member states. Furthermore, Börzel (2001) argues that the responsibility of EU law implementation is on the member states, which further weakens the European Commission since it is not directly politically legitimized (Börzel 2001, 812). Therefore, it may treat powerful member states like France, Germany, Italy, Spain and the Netherlands that contribute the most to the EU budget and that have much voting power in the Council more carefully than others (Börzel 2001, 812; Hix and Høyland 2011, 263–264). To control for the effect of these states, I create a dummy variable that takes the value of 1 for the countries Germany, France, Italy, Spain and the Netherlands and 0 for the remaining countries.

Control Variable 2 – Non–Compliance of other Member States: Increasing non–compliance with the SGP is likely to reduce voters' support for the Eurozone in other countries (Köhler et al. 2018, 4). Consequently, also support for the rules of the SGP as well as the Commission who is responsible to enforce these rules can be expected to decrease (Köhler et al. 2018, 14).

The Commission, in turn, reacts with increased enforcement in case of repeated non-compliance (Börzel 2003a, 198). In order to account for these two effects of non–compliance, a variable that counts the number of member states that violated the rules of the SGP in a given year will be included in the analysis.

Control Variable 3 – Eurozone Membership: Eurozone membership affects both public support and EDP initiations: first, Kuhn and Stöckel (2014) find that Eurozone member states are more in favor of European Economic Governance such as a strong coordination of economic and financial policies (Kuhn and Stöckel 2014, 630–635). Due to the important role of the SGP and the strong position of the Commission in economic policy coordination, we can expect that support for European Economic Governance directly translates into higher support for the SGP and the Commission in Eurozone countries (Bauer and Becker 2014, 219–222).

Second, even though the rules of the SGP apply to all members of the EU, the fiscal rules of the SGP are particularly relevant for the countries that are part of the currency area (Buti and Giudice 2002, 824–825; Faini 2006, 446–447; European Commission 2017d). This is because first, fiscal indiscipline of one member state can affect the overall Eurozone interest rate (and thus other members' fiscal situation) and second, price stability can no longer be guaranteed by the ECB in case of high deficits and debt (Buti and Giudice 2002, 824–825; Faini 2006, 446–447). For these reasons, the Commission might be more eager to sanction Eurozone members since excessive deficit or debt levels may threaten the stability of the common currency. Due to this twofold effect of Eurozone membership, I include a dummy variable that takes the value of 1 if a country is member of the Eurozone and 0 otherwise in the analysis.

Control Variable 4 – National Elections: National elections affect public support insofar that parties will more or less foster Euroscepticism among voters as a result of a new cleavage between winners and losers of European integration (Kriesi 2007, 85). In addition, national elections also influence the sanctioning behavior of the Commission. For example, in 2014 an 2015, the Commission refrained from enforcing compliance against

Spain due to upcoming elections (Wijsman and Crombez 2017, 5). In order to consider the effect of national elections, a dummy variable taking the value of 1 if either an executive or legislative election is taking place in a given year, is added to the model. The coding is based on two separate dummy variables that indicate the presence of either an executive or legislative election in a given year, respectively (Teorell et al. 2019, 179; Teorell et al. 2019, 184). A missing observation in the dummy variable on the occurrence of legislative elections for Estonia in 2016 was replaced by information on the occurrence of such an election provied by the OSCE (OSCE 2019).

While the above sections have described the dependent, independent and control variables in detail, Table 3 provides an overview of all variables included in the model as well as their expected effects. As for many social science research projects, the data for all of the indicators listed in the table include missing values. The next part will thus focus on the method of how I will deal with these missing observations.

Table 3: Overview: Variables, Operationalization and Expected Effect on EDP Initiation

Variable	**Operationalization**	**Effect**
Dependent Variable		
Excessive Deficit Procedure – Opinion	Dummy: no opinion on existence of excessive deficit = 0, opinion on existence of excessive deficit = 1	
Excessive Deficit Procedure – Report	Dummy: no initial Commission Report = 0, initial Commission Report = 1	
Independent Variables		
Absolute Level of Public Support (Vanberg 2005)	Share of respondents expressing trust in the European Commission in the Eurobarometer survey	H1: +
Relative Level of Public Support (Köhler et al. 2018)	Difference between share of respondents expressing trust in the European Commission and the share of respondents expressing trust in the national government in the Eurobarometer survey	H1: +
Transparency (Vanberg 2001)	Real GDP (at constant 2011 national prices in Millions, U.S. dollar)	H2: +
Absolute Level of Public Support × Transparency (Vanberg 2005)	Interaction term consisting of absolute public support indicator and Real GDP	H1: +
Relative Level of Public Support × Transparency (Vanberg 2005 and Köhler et al. 2018)	Interaction term consisting of relative public support indicator and Real GDP	H1: +

Control Variables		
Powerful Member States	Dummy: 1 = Germany, France, Italy, Spain and the Netherlands, 0 = for all other member states	
Non–Compliance of Other Member States	Number of Member States that violated the rules of the SGP per year	
Eurozone Membership	Dummy: 0 = no member of Eurozone, 1 = member of Eurozone	
National Elections	Dummy: 0 = no election, 1 = election	
Robustness Checks		
Positive Image of the European Union	Share of individuals that states to have a very or fairly positive image of the EU in the Eurobarometer survey	H1: +
Transparency	GDP Growth (%)	H2: +
Absolute Level of Public Support × Transparency	Interaction term consisting of absolute public support indicator and GDP Growth	H1: +
Relative Level of Public Support × Transparency	Interaction term consisting of relative public support indicator and GDP Growth	H1: +
Positive Image of European Union × Transparency	Interaction term consisting of Positive Image of the European Union and Real GDP	H1: +
Positive Image of European Union × Transparency	Interaction term consisting of Positive Image of the European Union and GDP Growth	H1: +

6.2.2 Data Imputation

Missing data is a well–known problem in Political Science research (King et al. 2001; König et al. 2005; Lall 2016). Unsurprisingly, many of the political and economic variables to be used for my analysis suffer from missing observations. Figure 11 [9] provides an overview of the amount of missing observations of each variable in the dataset [10].

[9]Figure 11 has been created using the `R`–package `naniar` (Tierney, Cook, McBain, and Fay Tierney et al.).

[10]Most missing values occurred for the Eurobarometer data for the countries Bulgaria, Cyprus, Czech Republic, Estonia, Croatia, Hungary, Lithuania, Latvia , Malta, Poland,

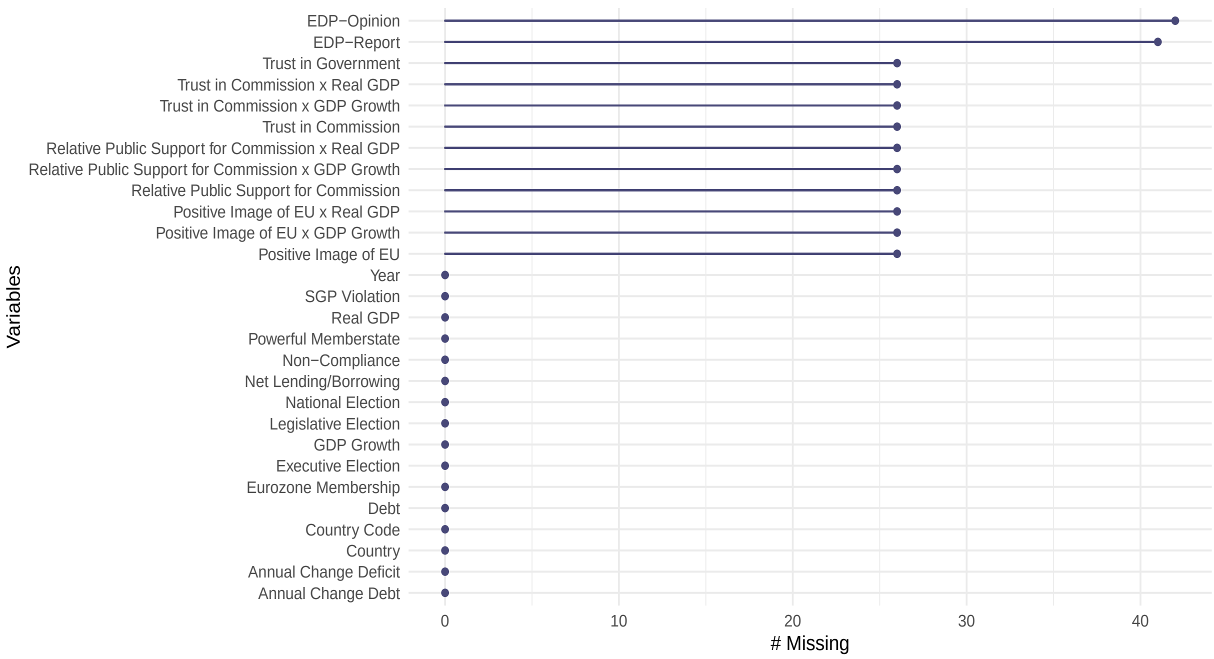


Figure 11: Overview of the Amount of Missing Observations per Variable

Usually, the most prominent method to deal with missing values is listwise deletion, namely the removal of missing observations from the dataset (Lall 2016, 414). Unfortunately, it has been confirmed that this method can bias the results of the following analysis (King et al. 2001; Lall 2016). The reason is that listwise deletion only yields unbiased results if the missing observations are *missing completely at random (MCAR)*, which means that the missingness of the observations does not depend on *any* observations in the dataset (König et al. 2005, 273; Lall 2016, 416). Since this assumption is not met in most of the cases, listwise deletion is problematic (König et al. 2005, 273).

Besides listwise deletion, there are other comparatively easy methods to deal with missing data. One of them is the imputation of missing values with the (unconditional) mean (Jerez et al. 2010, 107–108; König et al. 2005, 274; Lall 2016, 418; Little 988a, 288).

Romania, Slovakia and Slovenia for 2002 and 2003, as they were not part of the Eurobarometer Survey that includes the items on trust in the European Commission, Trust in the national government and the image of the EU (GESIS Leibniz Institute for the Social Sciences 2012a, VIII–IX; GESIS Leibniz Institute for the Social Sciences 2012b, VII–VIII; GESIS Leibniz Institute for the Social Sciences 2012c, VIII–IX). These missing values also contribute to the missing values in the indicator “Relative Public Support” and the interaction terms.

Furthermore, this tempting solution has a number of major drawbacks: For instance, mean imputation leads to the same (biased) estimate as listwise deletion (Little 988a, 288). In addition, standard errors are incorrect and downward–biased because they are a) calculated based on a overstated sample size and b) imputed values are treated as known rather than probabilistic estimates that account for uncertainty (Jerez et al. 2010, 107; Lall 2016, 418; Little 988a, 288). Second, because the same value is used for all cases with missing data, mean imputation can distort the variance of the variables with missing data (König et al. 2005, 274; Schafer and Graham 2002, 149). Third, the method can not be applied to variables measured at the nominal or ordinal level (König et al. 2005, 274). In sum, neither listwise deletion nor mean imputation can be viewed as appropriate methods to deal with missing data.

Due to the prominence of the missing value problem and the various disadvantages associated with simple solutions to it, more sophisticated methods have been developed. For these methods, the assumption that missing data are *missing at random (MAR)* – which means that the missingness depends on observed data only – needs to hold [11] (Lall 2016, 416–423).

Unfortunately, the MAR case is difficult to distinguish from data that are *missing not at random (MNAR)*, where the missingness depends (in part) on the missing data themselves (Lall 2016, 416–419). This identification is crucial, since even more sophisticated imputation methods will lead to biased results under MNAR (Lall 2016, 418). Unfortunately, it is not possible to test whether missing data are MAR or MNAR since the causes or correlates

[11] Little (1988b) proposes a test for the MCAR assumption, which I applied to the dataset used for this investigation using the `R`–package `BaylorEdPsych` (Beaujean 2012; Lall 2016, 421–423; Little 988b). The statistically significant result of the test confirmed that – as expected – the MCAR assumption does not hold, so data imputation is preferable to listwise deletion (Lall 2016, 421–423). For the test, the interaction terms, the dummy variables and the non–compliance indicator were excluded from the dataset due to high collinearity with the constitutive terms or the variables their coding was based on (Brambor et al. 2006, 70–71; Hardy 1993, 6). This is because the test yields no results in the presence of collinear variables since linear dependence in the data leads to a singular matrix where the linear system of equations has infinite solutions (Gill 2008, 146–159; Lall 2016, 423). Note that I do not follow Lall (2016) in assessing the MAR–assumption via logistic regressions on binary missingness indicators since it is generally impossible to test whether the assumption of MAR holds (Lall 2016, 418–423; Schafer and Graham 2002, 152).

of missing data are usually unknown to the researcher (Collins et al. 2001, 333). However, it is possible to apply imputation techniques that are based on MAR since deviations from this assumption may only have a minor impact on the results of the subsequent analysis (Collins et al. 2001, 333; Schafer and Graham 2002, 152).

Jerez et al. (2010) distinguish between statistical and machine learning data imputation methods (Jerez et al. 2010, 107). In Political Science, the statistical method of multiple imputation (MI) seems to be the most prominent advanced imputation method (King et al. 2001, 49; König et al. 2005, 274; Lall 2016, 417–418). MI is based on a *parametric* imputation model that assumes that the complete data can be described by a multivariate normal distribution (including unknown parameters) (Lall 2016, 418; Schafer 1999, 4–6). This usually means that missing data are be predicted based on selected variables in the dataset (Lall 2016, 425). MI is usually conducted via three stages: first, m values are imputed for each missing observation, which yields m complete datasets (King et al. 2001, 53; Lall 2016, 417–418). These imputed values are independent draws from a posterior distribution of the the missing data conditioning on data that can be observed (King et al. 2001, 53; Lall 2016, 418). The variation across these m imputed values indicates the uncertainty about the correct imputation model (King et al. 2001, 53; Lall 2016, 417–418).

In a second step, m quantities of interests are estimated from these datasets (King et al. 2001, 53; Lall 2016, 418). Third, the m estimates are combined to one average estimate (King et al. 2001, 53; Lall 2016, 418). The variance of this estimate is computed from the weighted sum of the estimated variances within and between the m datasets (King et al. 2001, 53; Lall 2016, 418). MI generally yields unbiased estimates in the case of MAR and MCAR (Jerez et al. 2010, 107; Lall 2016, 418; König et al. 2005, 270). Popular software packages are *Amelia* and *Amelia II* (for time–series cross–sectional data), which work with an Expectation Maximization algorithm (Honaker et al. 2011; King et al. 2001). Other methods include Bayesian approaches or even the consideration of strategic concealment (König et al. 2005, 285; Kong et al. 1994, 278).

Machine learning methods for data imputation seem so far to be less prominent in Political Science. Jerez et al. (2010) compare the performance of statistical imputation methods such as mean imputation, hot deck (= taking a missing observation from a similar donor case with complete data), and MI with machine learning methods (Multi–layer perceptron, Self–organization maps and k–nearest neighbors (KNN)) (Jerez et al. 2010, 105–108). Their analysis leads to two key findings: first, KNN, which imputes missing values with the mean (for continuous data) or the mode (for categorical data) of observed data points that are located closely to the missing value (defined by some distance measure), performs best (Jerez et al. 2010, 110–114). Second, machine learning methods generally outperform statistical methods of data imputation (Jerez et al. 2010, 113–114)[12].

Given the findings of Jerez et al. (2010), I choose a machine learning method instead of a statistical approach for missing data imputation. However, applying KNN is not possible here since my dataset includes both continuous and discrete variables (Stekhoven and Bühlmann 2012, 112).

As an alternative, Stekhoven and Bühlmann (2012) propose a Random Forest (RF)[13] approach, which outperforms KNN in data imputation and does not depend on variable type composition, data dimensionality or –complexity (such as spatio–temporal data) (Feng et al. 2014, 3592; Stekhoven and Bühlmann 2012, 113) .

In contrast to MI and KNN, RF is a *non–parametric* approach, which means that no specification of a parametric model or tuning parameters such as KNN's distance measure is necessary (Stekhoven and Bühlmann 2012, 112–114). This non–parametric approach is an advantage since selecting an inappropriate imputation model or the incorrect tuning parameter can strogly affect the performance of the imputation method (Stekhoven and Bühlmann 2012, 112).

According to Stekhoven and Bühlmann (2012), the RF algorithm for data imputation works as follows. At the beginning, the dataset is split into four parts for an arbitrarily variable X that includes missing values: first, the *observed* values of X; second, the *missing* values of X; and finally, the observed and missing values of the *remaining* variables besides X, respectively (Stekhoven and Bühlmann 2012, 113). As a first step, all variables in the dataset are sorted according to the number of missing values, where the starting point is the variable with the lowest amount of missing data (Stekhoven and Bühlmann 2012, 113). Thereafter, the missing values in the dataset are initially guessed; for example by

[12] Note that despite the application of a Bayesian model in the subsequent analysis, I do not apply a Bayesian imputation technique here since they rely on the Expectation Maximization (EM) algorithm, whose prognosis accuracy is found to be lower by Jerez et al. (2010) than the one of machine learning methods (Gelman et al. 2013, 454; Jerez et al. 2010, 111–113; King et al. 2001, 55; Kong et al. 1994, 278).

[13] A RF classifier is defined as a "[...] a classifier consisting of a collection of tree–structured classifiers $\{h(\mathbf{x}, \Theta_k), k = 1, ...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{x}$" (Breiman 2001, 6).

mean imputation or any another imputation method (Stekhoven and Bühlmann 2012, 113). Subsequently, an RF is fitted for each variable X using its observed observations as labels and the observed values of the other variables in the dataset as predictors (Stekhoven and Bühlmann 2012, 113). The missing values of X are then predicted by applying the trained RF to the missing observations of the remaining variables (Stekhoven and Bühlmann 2012, 113). Finally, the previously by initial guessing imputed dataset is updated by using the newly predicted values for the missings in each variable X (Stekhoven and Bühlmann 2012, 113). The algorithm stops when the difference between the updated (imputed) dataset and the roughly imputed dataset increases (Stekhoven and Bühlmann 2012, 113).

Since the RF approach fits with the structure of my dataset and does not require the (often incorrect) specification of an imputation model or any kind of tuning parameter, I use the `missForest` R–package provided by Stekhoven and Bühlmann (2012) for data imputation[14]

[14]`missForest` is – similar to Stekhoven and Bühlmann's (2012) experiments – applied to a dataset that includes transformed and untransformed variables like interaction terms and dummy variables (Stekhoven and Bühlmann 2012, 114–115). This procedure is in line with the recommendations of King et al. (2001), van Buuren and Groothuis-Oudshoorn (2011) and von Hippel (2009) to first transform, then impute (although the imputed values of the transformed variables can be very unintuitive) (King et al. 2001, 57; van Buuren and Groothuis-Oudshoorn 2011, 22; von Hippel 2009, 282–290). This is because they find that a) the exclusion of variables from imputation that are part of the statistical model and b) imputation of only nontransformed data biases regression results when a parametric imputation method is applied (King et al. 2001, 57; van Buuren and Groothuis-Oudshoorn 2011, 22; von Hippel 2009, 265–271). Despite the fact that `missForest` is a non-parametric imputation method, I will follow this advice since interactions and other (non-linear) transformations are considered by `missForest` and can thus be included in the imputation procedure to avoid biased estimates (Stekhoven and Bühlmann 2012, 112; Shah et al. 2014, 764; Doove et al. 2014, 93).

The imputation dataset further includes variables that are included in some regression models, but not in others (such as the variables used for the robustness checks). However, they are considered in the imputation process since observed values in some of these variables can be useful to predict missing values in other variables (King et al. 2001, 58). In addition, the dataset is reduced to cases of SGP violations for the analysis *after* imputation because the complete dataset is necessary for descriptive statistics that require variance

The RF algorithm and the `missForest`–package uses an *out–of–bag–error (OOB)*–estimate to determine imputation error (Breiman 2001, 11; Stekhoven and Bühlmann 2012, 112)[15]. The calculation of *OOB*–estimates differs for continuous and categorical variables (Stekhoven and Bühlmann 2012, 113). For continuous variables, the normalized root mean square error (NRMSE)[16] is provided, whereas the proportion of falsely classified entries (PFC) over the missing observations ($= \delta_F$) is computed for categorical variables (Stekhoven and Bühlmann 2012, 113–114). Generally, a good performance of RF yields a NRMSE/PFC value close to 0, whereas a value around 1 indicates a poor performance (Stekhoven and Bühlmann 2012, 114). For this project's dataset, the overall NRMSE is equal to 0.03, while the PFS is 0.00. This means that there is an imputation error of 3% for continuous variables and a 0% error for categorical variables, which indicates a good imputation performance of RF for this dataset. After the description of the data imputation process, I will now continue with the explanation of my statistical model.

---

between violations and non–violations, such as Table 1.

The only variables whose values are not imputed are both EDP measures, because for this variable, missing values only occur for countries that were not yet members of the EU at a given year and imputing non–existent values is not possible (King et al. 2001, 59). In order to ascertain that the imputation procedure did not lead to a bias in the analysis results by not conditioning on Y, I re-ran the analysis using complete cases because missing values in both EDP variables and Eurobarometer indicators occur for the same countries and years (von Hippel 2009, 266). As the absence of observations in both EDP variables does not depend on them itself – given that they are only due to non–EU membership at a certain point in time – listwise deletion does not introduce any bias in these analysis results in this special case (Van Buuren Van Buuren). However, this alternative procedure did not lead to any change in the results.

[15]For a further detailed explanation of the *OOB*–error estimate, see Breiman (2001).

[16]The NRMSE is defined by

$$NRMSE = \sqrt{\frac{((mean(X^{true} \smile X^{imp})^2)}{var(X^{true})}} \tag{6.1}$$

, where X^{true} is the complete and X^{imp} is the imputed data matrix (Stekhoven and Bühlmann 2012, 113–114).

7 Method: Binary Time–Series Cross–Sect Data Analysis

7.1 Characteristics of Binary Time–Series Cross–Sectional Data

The European Union comprises twenty–eight memberstates, while the Euro as a means of payment was introduced in 2002 and is currently used by nineteen memberstates (European Union 2018b). Nevertheless, the SGP applies to all EU states, not only the Eurozone members (European Union 2008). This investigation's dataset is an unbalanced time–series dataset that includes data of 25 countries and 16 years. Because I am interested in the determinants of sanctioning by the European Commission, the dependent variable is a binary variable taking the value of 1 if an EDP is initiated and 0 otherwise.

By consisting of observations for both different countries and years and the binary dependent variable, the data can be classified as BTSCS data (Beck et al. 1998, 1260–1262; Beck 2001a, 111–273; Beck 2001b, 271). According to Beck et al. (1998) and Beck (2001a; 2001b), BTSCS data can be described by

$$P(y_{i,t} = 1) = f(x_{i,t}\beta); \;\; i = 1...N; t = 1...T \tag{7.1}$$

where $y_{i,t}$ is the (observed) binary dependent variable, x_{it} is a vector of independent variables (both indexed by unit i and time t), and f is either a logistic or standard normal cumulative distribution function (cdf) (Beck et al. 1998, 1260–1262; Wooldridge 2002, 457–458; Wooldridge 2009, 575–576)[1].

[1]Note that there is a distinction between panel and (B)TSCS data: while panel data are sampled repeated cross–section data where units (usually countries) are only observed for a few times, (B)TSCS data consist of non–sampled, fixed units (Beck 2001a, 113; Beck 2001b,

Dealing with BTCS data bears several methodological challenges: First, the errors can be contemporaneously correlated, which means that the errors of two different units are associated for a given point in time (Beck and Katz 1995, 636). This problem can occur when – for example – the economies of two countries are linked or a common "time shock" affects all units in a given year (Beck and Katz 1995, 636; Shor et al. 2007, 166). Contemporaneous correlation in BTSCS data violates the assumption of independence of observations of logit or probit models (Beck et al. 1998, 1261–1263). Consequently, the application of these models to BTSCS data leads to incorrect standard errors and statistical tests as well as a loss of information in the data (Beck and Katz 1995, 636; Beck et al. 1998, 1261–1263; Beck 2001a, 128–130; Shor et al. 2007, 165–172)[2].

Second, the error variances of the units (usually countries) in BTSCS can differ, which is called "panel heteroskedasticity" or "unit heteroskedasticity" (Beck and Katz 1995, 636; Shor et al. 2007, 172). Panel heteroskedasticity in a linear model yields unbiased, but inefficient estimates (Yatchew and Griliches 1985, 135–137). By contrast, for a binary outcome models like a probit model, in constrast, heteroskedasticity is more problematic (Keele and Park 2005, 1–6; Yatchew and Griliches 1985, 135–138). To understand why that is, recall that the underlying assumption of a probit model is that the error term ϵ_{it} takes a standard normal distribution Φ with constant variance σ (which is equal to 1), which implies homoskedasticity (Greene 2012, 726; Keele and Park 2005, 6). Given this assumption, a probit model can be written as

$$P\{y_{it} = 1\} = \Phi(\frac{x_{it}\beta}{\sigma}) \tag{7.2}$$

From this term, it becomes obvious that if there is heteroskedasticity – i.e. σ is not constantly equal to 1 in a probit model – the estimate $\hat{\beta} = \frac{\hat{\beta}}{\sigma}$ is inconsistent and even biased if not appropriately modeled (Keele and Park 2005, 6; Yatchew and Griliches 1985, 135–138).

Third, the number of countries in TSCS/BTSCS data is rather small, usually 10 to 50

273). When using panel data, the main interest is in the whole population rather than in the observed sample (Beck 2001a, 113; Beck 2001b, 273). For (B)TSCS, in contrast, the focus is on the observed units, simply because resampling countries is not possible (Beck 2001a, 113; Beck 2001b, 273).

[2]Unfortunately, applying a formal test for cross–section dependence like the Breusch–Pagan Lagrangian Multiplier test is not possible for non–linear models since this test assumes a linear model (Hsiao et al. 2012, 254).

observations on 10 to 50 countries (Beck and Katz 2007, 183)[3]. Such small samples can be a problem for (frequentist) Maximum–Likelihood (ML) estimation since consistency, normality and efficiency of the ML estimator only hold if the sample size n increases up to infinity (King 1989, 77–83; Long 1997, 53–54). If this large n condition is not met, the sampling distribution of the ML estimator – which is required for the computation of confidence intervals – is unknown (Dixon 2002, 1; Long 1997, 53–54). Applying ML estimators on BTSCS data may therefore yield incorrect confidence intervals unless bootstrapping methods to simulate this unknown sampling distribution are applied (Dixon 2002, 1).

Over time, a number of different techniques have been proposed to analyze BTSCS data: on the one hand, the standard solution to the problem of contemporaneous correlation in BTSCS data is provided by Beck et al. (1998), who propose to add $k_{t^{\smile}t0}$ temporal dummies to the logit or probit model (Beck et al. 1998, 1267–1270; Leblang 1999, 606; Mansfield and Reinhardt 2003, 848; Pang 2010, 470–471). This approach can also be combined with robust standard errors to account for the problem of panel heteroskedasticity in BTSCS data (Beck et al. 1998, 1283). Unfortunately, their approach suffers from a number of drawbacks: first, robust standard errors are known to only correct model missspecification such as an omitted variable bias or fundamental heteroskedasticity (King and Roberts 2015, 159–165). Second, factors that determine country–level differences – which usually hold interest in BTSCS analyses – remain ignored (Beck 2001a, 113; Beck 2001b, 273). Even worse, accounting for country–level effects via adding country–fixed effects is not possible in non–linear models since this would lead to biased estimates due to the “incidental parameter problem”[4] (Greene 2012, 627–628; Lee 2016, 179–180; Wooldridge 2002,

[3]For example, Baerg and Hallerberg’s (2016) sample consists of 15 EU member states, Leblang (1999) focuses on 55 to 76 developing countries and Hallerberg (2002) tests his hypotheses on OECD countries only, which were less than 35 countries in the postwar period (Baerg and Hallerberg 2016, 979–980; Hallerberg 2002, 792; Leblang 1999, 606; OECD 2017).

[4]The “incidental parameter problem” occurs when T is fixed and $N \to \infty$ (Greene 2012, 659; Wooldridge 2002, 484–491). This is because the number of parameters to be estimated will increase with the sample size and their estimation is solely based on the T observations (Greene 2012, 659; Lee 2016, 179–180; Wooldridge 2002, 484). If T is fixed and small, the asymptotic variance of the maximum likelihood estimator does not decrease and results in inconsistent β parameters for a fixed effects probit model when using maximum likelihood estimation (Greene 2012, 659; Lee 2016, 179–180; Wooldridge 2002, 484).

484–491). Third, according to Carter and Signorino (2010), the approach often leads to inefficiency and separation since the inclusion of time dummies increases the number of parameters in the model and their values may perfectly be associated with values of either 0 or 1 of the dependent variable y [5] (Carter and Signorino 2010, 275–276).

On the other hand, multi–level models became increasingly popular to model (B)TSCS data (Beck and Katz 2007; Raffalovich and Chung 2014; Shor et al. 2007). This is because instead of modeling BTSCS as grouped duration data[6], they can be regarded as non–nested multi–level data where observations are nested within countries and within years (Beck et al. 1998, 1264–1270; Carter and Signorino 2010, 290; Gelman and Hill 2006, 241–244). Consequently, heterogeneity can be found on both the country and year dimension (Pang 2010, 473). In fact, both contemporaneous correlation and heteroskedasticity are inherent parts of multi–level models that are explicitly modeled, which makes them ideal for the application on (B)TSCS data (De Leeuw and Meijer 2008, 13–14; Pang 2010, 470–475; Shor et al. 2007, 165–173).

A number of studies have proposed the application of linear multi–level models on continuous TSCS data and assessed their performance compared with standard approaches (Beck and Katz 2007, Gelman and Hill 2006, Hedeker and Gibbons 2006, Fahrmeir et al. 2013, Raffalovich and Chung 2014, Shor et al. 2007). Unfortunately, modeling contemporaneous correlation is – with the exception of Shor et al. (2007) – usually ignored, despite the fact that it yields too small standard errors and confidence intervals, which eventually affects hypothesis testing (Beck et al. 1998, 1261–1264; Beck 2001a; Beck 2001b; Pang 2010, 488–489; Shor et al. 2007, 165–172). Furthermore, evidence is limited for BTSCS data, despite the severe consequences of unmodeled panel heteroskedasticity in non–linear

[5]An alternative technique to the Beck et al. (1998) approach proposed by Carter and Signorino (2010) is adding cubic polynomials t, t^2 and t^3 to the regression equation, which avoids the Beck et al. (1998) approaches' problems of inefficiency and complete separation (Carter and Signorino 2010, 271–278). However, their technique also only addresses the problem of temporal dependence, but not panel heteroskedasticity (Carter and Signorino 2010, 271–273).

[6]Beck et al. (1998) as well as Carter and Signorino (2010) see BTSCS data as grouped duration data where the hazard rate – namely, the likelihood of "failure" of a unit – is dependent on both independent variables and time, which makes it possible to account for temporal dependence in the regression model (Beck et al. 1998, 1264–1270; Carter and Signorino 2010, 290).

models (Keele and Park 2005, 6; Yatchew and Griliches 1985, 135–137).

One exception is Pang (2010) who proposes to model both heterogeneity and serial correlation in BTSCS data with a Bayesian multi–level probit model with a pth–order autoregressive $AR(p)$ error process and a country– and year–level intercept (Pang 2010, 470–475). One reason speaking for a Bayesian approach is that when dealing with BTSCS data in the EU/Eurozone context, the data already constitute of observations from a "population" of EU states, so the principle of classical frequentist statistics of drawing inferences from long–run repeated random sampling is inapplicable (Western and Jackman 1994, 412–416). Another reason is related to sample size: as already mentioned above, the dataset includes 25 countries only. In that context, Stegmüller (2013) has shown that for sample sizes of less than 30 countries, Bayesian multi–level probit models result in correct standard errors and confidence intervals compared with frequentist multi–level models, which usually require bootstrapping methods to properly compute these quantities in case of a small sample (Dixon 2002, 1; Stegmüller 2013, 755–758). However, despite the fact that Pang (2010) applies her model to some empirical examples, she does not provide a comparison of the performance of her model to the approach of Beck et al. (1998) or frequentist multi–level models in her Monte Carlo simulation (Pang 2010, 482–484). Therefore, the question of how well Bayesian BTSCS approaches perform compared with frequentist models for BTSCS data remains unanswered.

In order to overcome the challenges of analyzing BTSCS data in this project, I present a Bayesian multi–level probit model with a BTSCS specification [7] (from now onwards referred to as BMLP BTSCS model). Moreover, to answer the methodological question of which model is most suitable for small–sample BTSCS data analyses – such as in case of EU/Eurozone research – I compare the BMLP BTSCS' performance to five other models used for BTSCS data analysis in the Political Science literature – including Beck et al.'s (1998) BTSCS standard approach and frequentist multi–level probit models as recommended in standard multi–level textbooks – in a Monte Carlo experiment (Beck et al. 1998, 1261–1264; Fahrmeir et al. 2013, 350–391; Gelman and Hill 2006, 241–244; 289–290; Hedeker and Gibbons 2006; Pang 2010, 474–475; Shor et al. 2007, 170–172).

Because the BMLP BTSCS model a) accounts for both contemporaneous correlation and panel heteroskedasticity via its country– and year–level random intercepts and b)

[7] A BTSCS specification means that the model includes both a country– and a year–level random intercept to model both contemporaneous correlation and panel heteroskedasticity (Pang 2010, 470; Shor et al. 2007, 171–172).

Bayesian models have been found to outperform frequentist multi–level models for samples with few countries, I expect the BMLP BTSCS model to result in both unbiased estimates and correct standard errors for small samples compared with the other aforementioned BTSCS modeling techniques (Stegmüller 2013, 755–758; Pang 2010, 470; Shor et al. 2007, 171–172). The following section will discuss this model in further detail.

7.2 Modeling Binary Time–Series Cross–Sectional Data: A Bayesian Multi–Level Approach

The BTSCS structure of the data used here can also be regarded as non–nested multi–level data where observations are clustered in both $country \times year$ levels (Gelman and Hill 2006, 241–244). As for "nested" multi–level data structures (for example, where individuals are nested in groups), the variation of the levels (among countries and over years) is of interest for BTSCS data that can accordingly be modeled (Gelman and Hill 2006, 241–244; 289–290; Shor et al. 2007, 170–171). Due to the data's multi–level structure and the dichotomous nature of the dependent variable, I follow Stegmüller (2013) and choose a Bayesian multi–level probit model (Stegmüller 2013, 749).

The decision for a Bayesian approach stems from two advantages: from a methodological perspective, Bayesian models – in contrast to ML estimators – do not require bootstrapping to obtain correct standard errors and confidence intervals for small samples like the dataset used for this investigation (Dixon 2002, 1; Stegmüller 2013, 755–758). This is because the Bayesian approach is based on multiplying the likelihood with a prior to obtain the so–called "posterior distribution" (Gelman and Hill 2006, 392). From the posterior, inferences are then summarized from by random draws (Gelman and Hill 2006, 392). The posterior distribution is a real probability distribution, so Bayesian "credible intervals" and "standard errors" can be computed without referring to a hypothetical sampling distribution since they are simply the corresponding quantiles/standard deviation of that distribution (Lee et al. 2017, 868; Stegmüller 2013, 750).

More substantially, note that the data already include all of the available observations from a population of EU member states that have violated the SGP, so the application of frequentist methods that are based on a sampling notion is not very reasonable (Western and Jackman 1994, 412–415). By being based on the posterior probability distribution, the Bayesian approach does not rely on this notion: Bayesian credible intervals yield the posterior probability that a coefficient lies in that interval without any reference to a (hypothetical) population of countries (Stegmüller 2013, 750; Western and Jackman 1994, 412). Consequently, Bayesian models are more reasonable for data constituting of countries that – for example – belong to certain unique organizations (Western and Jackman 1994, 412).

I will apply a probit instead of a logit model due to the differences in data augmentation between these two models (Jackman 2009, 393). Given the conditional independence of the data, in line with Jackman (2009) the likelihood for the parameter estimates β for a binary response model can be described by

$$\mathcal{L}(\beta; y, X) = p(y|X, \beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}[1\breve{}F(x_i\beta)]^{1\breve{}y_i} \tag{7.3}$$

where for $y_i = 1$, the likelihood is $\pi_i = F(x_i\beta)$, whereas for $y_i = 0$, it equals $\pi_i = 1\breve{}F(x_i\beta)$ (Jackman 2009, 380). The problem for binary response models is that, given this likelihood, there is no conjugate prior for β (Jackman 2009, 380). However, the probit model can be turned into a linear regression conditional on latent data, a process called *data augmentation* (Jackman 2009, 380). Via this process, running a Gibbs sampler to conduct Bayesian analysis is feasible (Jackman 2009, 380). Data augmentation is based on the latent variable approach of binary response models (Jackman 2009, 380). For example, a probit model can be written as a latent variable regression model with normal errors and a censored latent dependent variable y_i^*:

$$y_i^* = x_i\beta + \epsilon_i, \;\; \epsilon_i \sim \mathcal{N}(0, 1) \; \forall i = 1, ..., n \tag{7.4}$$

with the censoring rule $y_i = 0 \Rightarrow y_i^* < \tau$ and $y_i^* = 1 \Rightarrow \geq \tau$ (Jackman 2009, 380).

Due to the latent data $y^* = (y_1^*, ...y_n^*)$, it is possible to work with the “data augmented posterior density”, $p(\beta|y^*, y, X)$, instead of computing the posterior $p(\beta|y, X)$ directly (which would be much harder for the Gibbs sampler) (Jackman 2009, 381). Sampling from the posterior for β via an augmented Gibbs sampler proceeds as follows: first, the augmented data y^* are generated from the predicted density $p(y^*|\beta, y, X)$ (Jackman 2009, 381). Subsequently, β is sampled from its conditional density given the augmented data and observed data $p(\beta|y^*, y, X)$ (Jackman 2009, 381). For the probit model, the conditional density of β equals $p(\beta|y^*, X)$ since conditioning on y is not necessary given y^* (Jackman 2009, 381).

This process is easier to implement for the probit model for two reasons: first, the normal density underlying the probit model implies that

$$y_i^*|x_i, \tilde{\beta}, \;\; y_i = 0 \sim \mathcal{N}(x_i\tilde{\beta}, 1)\mathcal{I}(y_i^* < 0) \tag{7.5}$$

$$y_i^*|x_i, \tilde{\beta}, \;\; y_i = 1 \sim \mathcal{N}(x_i\tilde{\beta}, 1)\mathcal{I}(y_i^* > 0) \tag{7.6}$$

where $\mathcal{I}(\cdot)$ is an indicator function which takes the value of 1 if its argument is true and zero otherwise, in order to induce truncated normal sampling (Jackman 2009, 381). Second, since the model for the latent variable y_i^* is a normal linear regression model, it is possible to use normal priors for β, $\beta \sim \mathcal{N}(b_0, B_0)$ because it is the conjugate prior for the normal model (Jackman 2009, 381). Furthermore, this enables one to use standard results on the Bayesian analysis of a linear regression problem (Jackman 2009, 381).

However, in case of a logit model, the latent variable y_i^* has logistic densities with mean $x_i\beta$ and is truncated given the observed y_i (Jackman 2009, 393). In contrast to the probit model, the problem here is that there is no corresponding conjugate prior for β (Jackman 2009, 393). Using a normal prior for β while having a logistic density for y_i^* leads to a conditional distribution for β, $p(\beta|y_i^*, X)$, which is no standard distribution (Jackman 2009, 393). Because it is more difficult for the Gibbs sampler to sample from this distribution (although there are solutions for this problem, see Jackman 2009), I rely on the multi–level probit model due to its smoother implementation.

Since a multi–level approach makes it possible to model the variation of the country– and year–levels, I will apply a Bayesian multi–level probit model with a BTSCS specification (henceforth referred to as BMLP BTSCS model) for the analysis. My model is a modification of BTSCS approaches presented by Pang (2010)[8] and Shor et al. (2007)[9], but deals with the problems that are associated with panel heteroskedasticity in N, contemporaneously–correlated errors in T and a small N simultaneously by including both a country– and year–specific random intercept and the application of a Bayesian approach (Pang 2010, 470–475; Shor et al. 2007, 170–172; Stegmüller 2013, 755–758). Thus, the BMLP BTSCS should result in both unbiased estimates and correct standard errors even for samples with few countries.

[8]Pang (2010) presents a Bayesian multi–level probit model that accounts for country– and year–level heterogeneity but also serial correlation in the individual errors (Pang 2010, 470–475).

[9]Shor et al. (2007) present a Bayesian multi–level linear regression model that models both panel heteroskedasticity and contemporaneous correlation (Shor et al. 2007, 165–172).

Using the latent variable formulation of Greenberg (2008) and McCulloch (2003) and the error term formulation of Gelman and Hill (2006), my model can be described by

$$y^*_{i,t} = \mu + X_{it}\beta + \alpha_i + \gamma_t + \epsilon_{it}; \;\; i = 1...N; \;\; t = 1...T \tag{7.7}$$

where $y^*_{i,t}$ is an underlying latent variable indexed by country i and year t; α_i and γ_t are the country– and year–specific random error term parameters, respectively, with

$$\alpha_i \sim \mathcal{N}(0, \sigma^2_\alpha) \tag{7.8}$$

$$\gamma_t \sim \mathcal{N}(0, \sigma^2_\gamma); \tag{7.9}$$

(Gelman and Hill 2006, 263–265; McCulloch 2003, 2–4; Rabe-Hesketh et al. 2005, 304; Shor et al. 2007, 170). Note that if the model has an overall intercept besides a country– and year–level error, the means of the latter have to be equal to 0 to make the model identifiable (Gelman and Hill 2006, 380–381; Pang 2010, 475) [10]. Moreover, α_i and γ_t are independent from ϵ_{it} and x_{it} (Rabe-Hesketh et al. 2005, 304). Finally, the model can be written as GLMM by

$$P\{y_{it} = 1|\alpha, \gamma\} = P\{X_{it}\beta + \alpha_i + \gamma_t + \epsilon_{it} > 0\} \tag{7.10}$$

$$P\{y_{it} = 1|\alpha, \gamma\} = \Phi\{X_{it}\beta + \alpha_i + \gamma_t\} \tag{7.11}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution link function that makes the model appropriate for binary outcome variables (Gelman and Hill 2006, 264–265; McCulloch 2003, 4).

The TSCS component of the model is the time–varying error parameter γ_t (Goldstein 1986; Shor et al. 2007, 170). Including γ_t makes it possible to account for contemporaneous correlation in the form of "time shocks" that affect all countries in a given year (Shor et al. 2007, 171–172). Moreover, the mean of α_i and γ_t is set equal to 0 since the constant term has been included as μ–parameter in the model (Gelman and Hill 2006, 381)[11]. Finally, β

[10]Note that without the baseline constant, the probabilities predicted by the model only depend on the relative values of the parameters (Gelman and Hill 2006, 315–316).

[11]This is because for non–nested models, one can either include a constant term in the model while setting the mean of α_i and γ_t to zero or define a mean parameter for

is a (non–random) coefficient with a single independent variable x_{it} that varies across time and units, and ϵ_{it} is the error term of level one (Shor et al. 2007, 170–171).

From equation 8 to 12, one can see that the BMLP BTSCS considers both panel heteroskedasticity and contemporaneous correlation. Therefore, it should thus yield unbiased estimates and correct credible intervals. Due to the Bayesian approach, this should also be the case for samples with a small number of countries (Stegmüller 2013, 755–758).

Instead of relying on repeatable experiments, the Bayesian approach multiplies the likelihood function – which is defined as the probability of the data given the parameters and independent variables – with a prior distribution (Gelman and Hill 2006, 388–392; Western and Jackman 1994, 412–416). This prior distribution represents non–sample prior information and needs to be selected for all parameters included in the model (Stegmüller 2013, 750; Western and Jackman 1994, 415). In multi–level models, priors for both the β coefficient– and the μ constant term – parameters as well at the hyperparameters σ_η and σ_ν need to be selected (Gelman and Hill 2006, 346).

For the selection of the β and μ parameter priors, I follow Gelman and Hill (2006), Martin (2008) and Pang (2010) and choose so called "noninformative" priors like $\beta \sim (0, 0.001)$ and $\mu \sim (0, 0.001)$ (Gelman and Hill 2006, 347–380; Martin 2008, 507; Pang 2010, 482; Shor et al. 2007, 170; Stegmüller 2013, 751).

In contrast to the linear model as proposed by Shor et al. (2007), choosing a non–informative uniform prior for α_i and γ_t is not possible in a non–linear model since this would make the α_i/γ_t–parameters equivalent to to frequentist fixed effects, which are known to suffer from the "incidental parameter problem" that yields biased estimates in non–linear models (Hahn 2004, 137; Lee 2016, 181). However, when applying a Bayesian approach with non–uniform priors, the incidental parameter problem is no longer present since the unit–varying random effects – in constrast to the frequentist framework – are estimated as a random variable (Lee 2016, 180–181).

For the choice of a non–uniform prior, I follow Lee (2016) and Gelman (2006), I choose $p(\sigma_\alpha) \sim 1/\sigma_\alpha$ and $p(\sigma_\gamma) \sim 1/\sigma_\gamma$ as hyperparameter priors which avoids making α_i and γ_t fixed effects and therefore the incidental parameter problem (Gelman 2006, 521; Lee 2016, 181). Accordingly, the Bayesian multi–level probit is equivalent to the frequentist random effects model and results in unbiased parameter estimates (Lee 2016, 181–184).

---

the country– and year–specific random effects without a constant since the inclusion of a constant term in more than one place leads to nonidentifiability (Gelman and Hill 2006, 381).

7.3 Monte Carlo Experiment

Thus far, the superiority of the BMLP BTSCS model is merely a theoretical expectation since there is surprisingly little empirical evidence of this in the methodological literature. Therefore, I test the performance of the BMLP BTSCS model compared with the performance of five other models used for BTSCS data analysis, which shall be described in further detail in the following section.

7.3.1 Models to be compared

The Monte Carlo experiment will compare six models overall in terms of bias and correct standard errors since panel heteroskedasticity and contemporaneous correlation are known to affect these two quantities in a probit model. Besides the BMLP BTSCS model, five additional models are considered in the model comparison[12]:

The first model is a frequentist pooled probit model (referred to as *Probit Model* in the experiment) that ignores both panel heteroskedasticity and contemporaneous correlation (as used by Hallerberg 2002) (Hallerberg 2002, 798). I expect the estimates of this model to be biased and standard errors/confidence intervals to be too small.

The second model is Beck et al.'s (1998) BTSCS time dummy approach (labeled as *B & K* here) that neglects country–level differences[13] (Beck et al. 1998, 1264–1270). Due to its ignorance of panel heteroskedasticity, I expect its estimates to be biased, but standard errors/confidence intervals to be correct.

The third and fourth model are both a Bayesian (*BMLP*) and a frequentist multi–level probit model (*Freq. MLP*) with a country–level intercept that both ignore contemporaneous correlation, as suggested by Fahrmeir et al. (2013) (Fahrmeir et al. 2013, 389–391; Pang 2010, 488). As such, I expect their estimates to be unbiased, but standard errors/-confidence intervals and Bayesian credible intervals to be too small. Given the findings

[12]The Bayesian models were run using the `R`–packages `coda` and `R2jags`, while the package `lme4` was used for the frequentist multi–level models (Bates et al. 2015); Plummer et al. 2006; Su and Yajima 2015).

[13]The approach of Carter and Signorino (2010) is not considered in the experiment since the time–dummy approach of Beck et al. (1998) is still the most frequently used one (2729 vs. 1081 Google Scholar citations (Google 2018)). The spline approach of Beck et al. (1998) is not part of the model comparison because it is relatively unknown among political scientists (Carter and Signorino 2010, 278).

of Stegmüller (2013), I particularly expect standard errors and confidence intervals of the frequentist model to be smaller than the credible intervals of the Bayesian model for a low number of countries (< 30) (Stegmüller 2013, 755–758).

The fifth model is a frequentist multi–level probit model with a country– and year–level intercept (as applied by Baerg and Hallerberg (2016) and denoted as *Freq. MLP BTSCS* here) to assess the difference between classical and Bayesian BTSCS models for small sample sizes. While this model accounts for both panel heteroskedasticity and contemporaneous correlation, I expect its estimates to be unbiased, but credible intervals to be small for less than 30 countries[14] (Stegmüller 2013, 755–758.).

The last and most important model included for comparison is the Bayesian multi–level probit model with a country– and year–specific random intercept (*BMLP BTSCS*). Since the BMLP BTSCS – as explained above – simultaneously accounts for panel heteroskedasticity and contemporaneous correlation and the Bayesian approach is suitable for samples including few countries, it should result in unbiased estimates and correct credible intervals[15] (Stegmüller 2013, 755–758).

7.3.2 Data Generating Process

In order to test whether the BMLP BTSCS model captures the peculiarities of BTSCS data better than other techniques, I will set up data with a variance–covariance matrix that is panel heteroskedastic and contemporaneously correlated (Beck and Katz 1995, 636; Beck et al. 1998, 1261–1263; Pang 2010, 472–475; Shor et al. 2007, 171–173).

Using the error component notation of Wallace and Hussain (1969)[16], the data gener-

[14] Due to convergence issues, the continuous independent variables have been scaled (Long 1997, 59–60). This usually solves the problem of imprecise gradient estimates (Long 1997, 59–60).

[15] All Bayesian models were both run with three chains iterating 8000 times where half of the interations where discarded as a burn–in and the thinning interval was set equal to 4, leaving 1000 iterations saved (Gelman et al. 2013, 282–283; Shor et al. 2007, 173). Chain covergence was confirmed for all parameters via the Gelman–Rubin $\hat{R}$ provided by `R`'s `coda`–package (Cowles and Carlin 1996, 884–885; Gelman et al. 2013, 284–285; Plummer et al. 2006; Shor et al. 2007, 173).

[16] Note that this notation is equivalent to the varying–coefficient notation previously used in this chapter (Gelman and Hill 2006, 262–265). However, I use the error component notation here since it simplifies the derivation of the data generation processes' covariance

ating process of the Monte Carlo experiment can be described as follows:

$$y^*_{it} = \mu + \beta x_{it} + \epsilon_{it}; \;\; i = 1...N; \;\; t = 1...T \tag{7.12}$$

$$\epsilon_{it} = \alpha_i + \gamma_t + \omega_{it} \tag{7.13}$$

$$\alpha_i \sim \mathcal{N}(0, \sigma^2_{\alpha_i}); \tag{7.14}$$

$$\gamma_t \sim \mathcal{N}(0, \sigma^2_{\gamma}); \tag{7.15}$$

$$\omega_{it} \sim \mathcal{N}(0, \sigma^2_{\omega}); \tag{7.16}$$

where $y^*_{i,t}$ is an underlying latent variable indexed by country i and year t; μ is the overall intercept set equal to 1, β is some regression parameter of the time– and unit–varying independent variable x_{it} which takes the value of 0.3 in the experiment [17], and ϵ_{it} is the composite error term (Greenberg 2008, 123; McCulloch 2003, 4; Shor et al. 2007, 170; Singer and Willett 2003, 247; Wallace and Hussain 1969, 56–58). Here, α_i accounts for the country–level errors, γ_t for the year–level errors and ω_{it} for the (individual) country–year–level 1 errors for which the following assumptions are imposed (Shor et al. 2007, 171; Singer and Willett 2003, 244–247; Wallace and Hussain 1969, 56–58):

1. $E[\alpha_i \alpha_{i'}] = 0$ for $i \neq i'$, namely the (country–level) errors of different countries are uncorrelated (Wallace and Hussain 1969, 56–58).

2. $[E\gamma_t\gamma_{t'}] = 0$ for $t \neq t'$, namely the (year–level) errors of different years are uncorrelated (Wallace and Hussain 1969, 56–58).

3. $[E\omega_{it}\omega_{i't}] = E[\omega_{it'}\omega_{it}] = [E\omega_{i't'}\omega_{it}] = 0$ for $i \neq i'$ and $t \neq t'$, namely the (level 1) errors are uncorrelated for different countries within a given year, for different years within a given country and for different countries and different years (Wallace and Hussain 1969, 56–58).

matrix.

[17]In line with Gelman and Hill (2006), I did not set the parameter values to either 0 or 1 to avoid masking programming errors (Gelman and Hill 2006, 363–364).

Recall that panel heteroskedasticity implies a constant variance over time *within* countries, but a different one *between* countries (Beck and Katz 1995, 636; Shor et al. 2007, 172). Thus, in order to induce panel heteroskedasticity in the experiment, the country–level error variances σ^2_α will be manipulated by varying with the independent variable x_{it} and some constant ψ to induce heteroskedasticity such that $\sigma^2_{\alpha_i}$ can be defined by

$$\sigma^2_{\alpha_i} = e^{x_{it} \cdot \psi}, \;\; i = 1, 2, 3...N; \tag{7.17}$$

(Beck and Katz 1995, 636; Carsey and Harden 2013, 107–108; Shor et al. 2007, 172–173; Singer and Willett 2003, 249–250)[18]. For the experiment conducted here, ψ is set to 0.8, whereby larger values of the independent variable result in larger country–level error variances (Carsey and Harden 2013, 107–108). Manipulating σ^2_α also leads to different levels of within–country autocorrelation of errors of *different years* across countries (Shor et al. 2007, 171; Singer and Willett 2003, 249).

Since the outcome variable needs to only take binary values based on the value of the latent variable y^*_{it} and the threshold 0, equation (8) is plugged in a standard normal cumulative distribution function yielding

$$P\{y_{it} = 1\} = P\{x_{it}\beta + \epsilon_{it} > 0\} \tag{7.18}$$

$$P\{y_{it} = 1\} = \Phi\{x_{it}\beta\} \tag{7.19}$$

(Gelman and Hill 2006, 264–265; McCulloch 2003, 4; Pang 2010, 474–475).

The crucial element to illustrate the characteristics of the BTSCS data created by the above described data generating process is the $NT \times NT$ covariance matrix of the composite errors ϵ_{it}, where N equals the number of countries and T the number of years (Wallace and Hussain 1969, 56–58). This covariance matrix is defined as

$$\sum = \sigma^2_\omega I_{NT} + \sigma^2_{\alpha_i} D + \sigma^2_\gamma B \tag{7.20}$$

where $\sigma^2_\omega I_{NT}$ is the level 1–error variance component, $\sigma^2_{\alpha_i} D$ the (heteroskedastic) country–level error variance component and $\sigma^2_\gamma B$ the year–level error variance component (Beck and Katz 1995, 636; Carsey and Harden 2013, 107–108; Shor et al. 2007, 172–173; Singer

[18]Using the exponential function in Equation (13) prevents negative variance values (Carsey and Harden 2013, 107–108).

and Willett 2003, 249–250; Wallace and Hussain 1969, 56–58). In order to illustrate how panel heteroskedasticity and contemporaneous correlation are included in $\sum$, let $N = 3$ (= three countries) and $T = 3$ (= three years). The covariance matrix $\sum$ is a 9×9 matrix of the following structure[19]:

$$\sum = \begin{bmatrix}
\sigma^2_\omega + \sigma^2_{\alpha_1} + \sigma^2_\gamma & \sigma^2_{\alpha_1} & \sigma^2_{\alpha_1} & \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & 0 & 0 \\
\sigma^2_{\alpha_1} & \sigma^2_\omega + \sigma^2_{\alpha_1} + \sigma^2_\gamma & \sigma^2_{\alpha_1} & 0 & \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & 0 \\
\sigma^2_{\alpha_1} & \sigma^2_{\alpha_1} & \sigma^2_\omega + \sigma^2_{\alpha_1} + \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma \\
\sigma^2_\gamma & 0 & 0 & \sigma^2_\omega + \sigma^2_{\alpha_2} + \sigma^2_\gamma & \sigma^2_{\alpha_2} & \sigma^2_{\alpha_2} & \sigma^2_\gamma & 0 & 0 \\
0 & \sigma^2_\gamma & 0 & \sigma^2_{\alpha_2} & \sigma^2_\omega + \sigma^2_{\alpha_2} + \sigma^2_\gamma & \sigma^2_{\alpha_2} & 0 & \sigma^2_\gamma & 0 \\
0 & 0 & \sigma^2_\gamma & \sigma^2_{\alpha_2} & \sigma^2_{\alpha_2} & \sigma^2_\omega + \sigma^2_{\alpha_2} + \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma \\
\sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & 0 & 0 & \sigma^2_\omega + \sigma^2_{\alpha_3} + \sigma^2_\gamma & \sigma^2_{\alpha_3} & \sigma^2_{\alpha_3} \\
0 & \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & 0 & \sigma^2_{\alpha_3} & \sigma^2_\omega + \sigma^2_{\alpha_3} + \sigma^2_\gamma & \sigma^2_{\alpha_3} \\
0 & 0 & \sigma^2_\gamma & 0 & 0 & \sigma^2_\gamma & \sigma^2_{\alpha_3} & \sigma^2_{\alpha_3} & \sigma^2_\omega + \sigma^2_{\alpha_3} + \sigma^2_\gamma
\end{bmatrix}$$

Note that $\sum$ matrix has the following properties:

— The block submatrices on the main diagonal of $\sum$ reflect the within–country variance– and covariance of the residuals for *different years within the same country* (Singer and Willett 2003, 249–750). The $\sigma^2_\omega + \sigma^2_{\alpha_i} + \sigma^2_\gamma$ term on the diagonals of these submatrices is the correlation of errors for the *same country* and the *same year* and indicate the *within*–country–variance over time, comprising the level 1–error–, country–error – and year–error variance parameters (Singer and Willett 2003, 249). It is important to mention here that the addition of the σ^2_γ year–error variance parameter is a key difference between the compound–symmetric covariance matrix design of a classical random–intercept model where occasions/years (level 1) are seen as nested in units (countries) and the covariance matrix of (B)TSCS data (Hedeker and Gibbons 2006, 49–52) [20]. The non–zero off–diagonals of the submatrices, denoted with $\sigma^2_{\alpha_i}$, indicate the amount of autocorrelation of errors within a given country that varies across countries (Shor et al. 2007, 171; Singer and Willett 2003, 249–250).

— Due to the manipulation described above, the $\sigma^2_{\alpha_i}$ country–variance parameters are unequal between units (which is indicated by the subscripts $i = 1, 2, 3...N$), thereby simulating panel heteroskedasticity in BTSCS data (Beck and Katz 1995, 636; Shor et al. 2007, 172–173).

[19]For detailed calculation, see Appendix.

[20]For the classical random intercept model, the diagonals of the submatrices are equal to $\sigma^2_\omega + \sigma^2_\alpha$ (Hedeker and Gibbons 2006, 49–52).

— The σ_γ^2 parameter on the main diagonal of the off–diagonal block matrices of $\sum$ is the correlation of errors for *different countries* and the *same year*, which reflects the notion of "contemporaneous correlation" in (B)TSCS data (Beck and Katz 1995, 636; Beck et al. 1998, 1261–1263; Pang 2010, 471–475; Shor et al. 2007, 172). In contrast to the country–level error variances $\sigma_{\alpha_i}^2$, σ_γ^2 and thus the level of contemporaneous correlation is assumed to be constant across both time and countries (Beck and Katz 1995, 636). In the experiment, σ_γ^2 takes the value of 0.7, whereby a strong level of contemporaneous correlation is simulated[21].

— Finally, the 0's indicate the assumed zero correlation of errors for both *different countries* ***and*** *different years* (Wallace and Hussain 1969, 56–58). For this experiment, σ_ω^2 is set equal to 1 (Rabe-Hesketh et al. 2005, 304).

Since BTSCS data typically comprise 10 to 50 observations on between 10 and 50 units, both N and T vary between 10 and 50 in the experiment (Beck and Katz 2007, 183). Consequently, the dimension of the $NT \times NT$ covariance matrix of the composite errors will vary accordingly (Wallace and Hussain 1969, 56–58). This allows me to assess the performance of the BMLP BTSCS compared with frequentist models for a small number of countries ($N < 30$) while the errors of T are contemporaneously correlated (Beck and Katz 1995, 636; Beck et al. 1998, 1261–1263; Maas and Hox 2004, 127–130; Pang 2010, 471–475; Shor et al. 2007, 172; Stegmüller 2013, 748–751). Note that this is an extension of Stegmüller's (2013) experiment which only focused on variation in N and intraclass correlation (Stegmüller 2013, 751).

In sum, $\sum$ reflects both panel heteroskedasticity and contemporaneous correlation while N and T vary between 10 and 50, so that the simulated data possess all three key characteristics of (B)TSCS data (Beck and Katz 1995, 635–636; Beck et al. 1998, 1260–1263; Shor et al. 2007, 171–173; Singer and Willett 2003, 249; Pang 2010, 475). For the experiment, I follow Shor et al. (2007) and perform 250 Monte Carlo trials (instead of 1,000 or more) due to the time consuming Bayesian estimators (Shor et al. 2007, 173).

[21]This scenario is particularly realistic for economies that are closely located to each other – such as within the European Union – where economic shocks in a given year can simultaneously affect neighboring countries (Beck and Katz 1995, 636).

7.3.3 Results

In order to assess each models' performance in terms of bias and standard errors, I computed several test statistics such as bias, mean–squared error and standard errors of the β parameters (Carsey and Harden 2013, 97–102; Shor et al. 2007, 174–176)[22]. In addition, I present a graphical illustration of the β–coefficients and their confidence/credible intervals.

Absolute Bias

Ignoring panel heteroskedasticity in BTSCS data can lead to incorrect estimates in case of non–linear models (Yatchew and Griliches 1985, 135–137). To assess each model's performance on dealing with that problem, I computed the absolute bias of the coefficients as $|\hat{\beta}\breve{}\beta|$ as a first test statistic (Carsey and Harden 2013, 97). Consequently, a probit model that considers panel heteroskedasticity should yield a bias close or equal to 0, while a model that ignores this problem should exhibit a larger bias. Figure 12 and 13[23] show each model's average bias variation over N and T, respectively (Carsey and Harden 2013, 97; Shor et al. 2007, 174–176)[24].

[22]Note that the Beck et al. (1998) approach suffered from the problem of separation in the Monte Carlo analysis which is likely to cause estimation problems, as already pointed out by Carter and Signorino (2010) (Carter and Signorino 2010, 271–278).

[23]The plots were created using the `R`–packages `dplyr`, `gdata` and `ggplot2` (Warnes et al. 2017; Wickham 2016; Wickham et al. 2017).

[24]Table 6 (see Appendix) shows the absolute bias of each model's estimates for each value of N and T. For the creation of the table, the `R`–packages `dplyr`; `reshape2`; `splitstackshape` and `xtable` were used (Dahl 2016; Mahto 2018; Wickham 2007; Wickham et al. 2017).

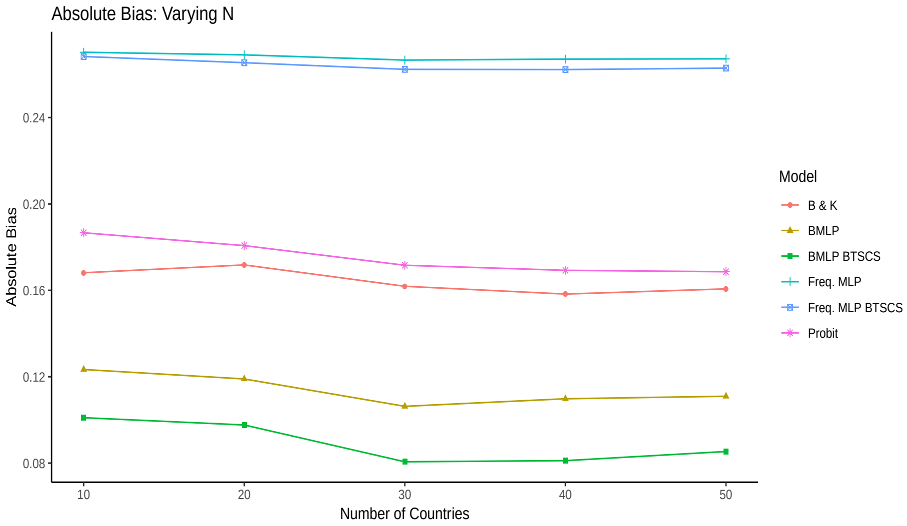


Figure 12: Average Absolute Bias of β–Coefficients for N Ranging from Minimum to Maximum

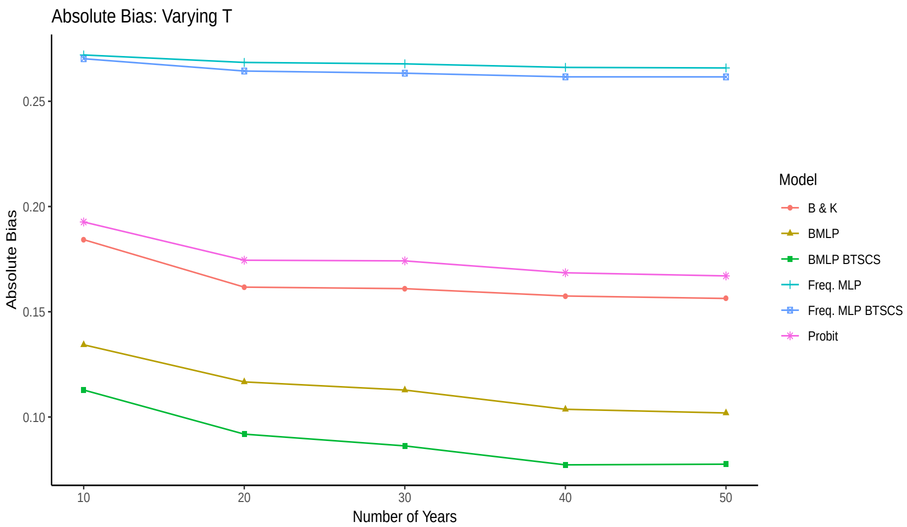


Figure 13: Average Absolute Bias of β–Coefficients for T Ranging from Minimum to Maximum

Although all models' estimates have a fairly small bias (see Table 6 in the Appendix), the BMLP BTSCS model consistently yields the smallest average bias, regardless of the size of N or T. It is only slightly larger for very small samples where both N and T are equal to 10. This indicates that the model deals appropriately with the problem of panel heteroskedasticity compared with alternative techniques. However, the absolute bias does not allow one to draw any conclusions about the efficiency of a parameter estimate (Carsey and Harden 2013, 98–136).

Mean Squared Error

In contrast to the absolute bias, the Mean Squared Error is both a measure of efficiency (= variability of the parameter estimate across the Monte Carlo trials) and bias since it puts more emphasis on variability (Carsey and Harden 2013, 98–136). The MSE of a parameter estimate is equal to the expectation of its squared deviation from the true parameter value (which is equal to 0.3 in this experiment), namely $MSE = E[(\hat{\beta}\,\breve{}\,\beta)^2]$ (Carsey and Harden 2013, 98–99). As for the absolute bias, we want the MSE to be as small as possible for a well–performing model (Carsey and Harden 2013, 98–99). The average MSEs for N and T of each of the models to be compared in this experiment are illustrated in Figure 14 and 15 [25].

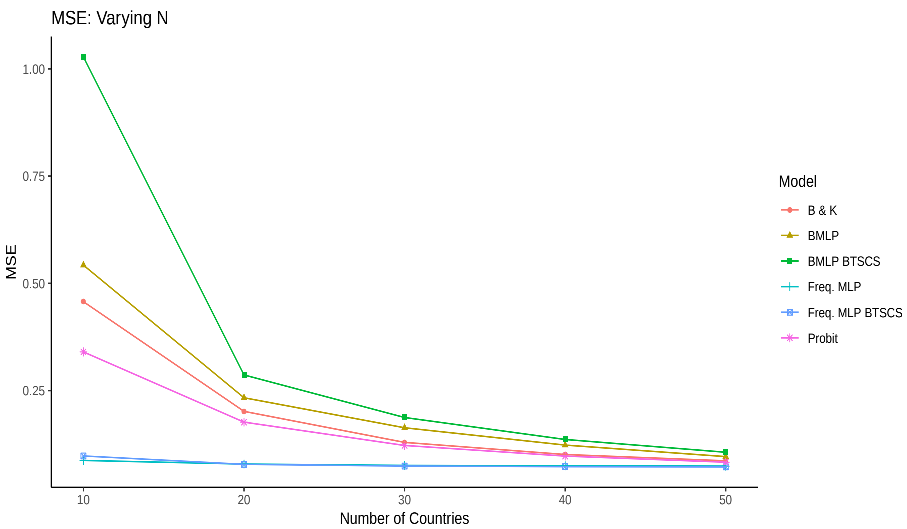


Figure 14: Average MSE of β–Coefficients for N Ranging from Minimum to Maximum

[25] The graphs were created using the R–packages `dplyr`, `gdata` and `ggplot2` (Warnes et al. 2017; Wickham 2016; Wickham et al. 2017). In addition, Table 7 (see Appendix) shows the MSE of each model's estimates for each value of N and T. The table was created using the R–packages `dplyr`; `reshape2`; `splitstackshape` and `xtable` were used (Dahl 2016; Mahto 2018; Wickham 2007; Wickham et al. 2017).

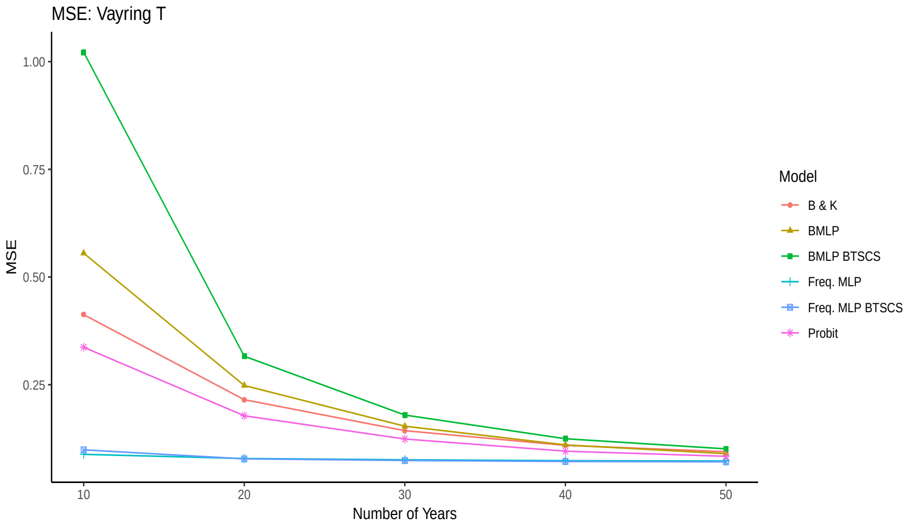


Figure 15: Average MSE of β–Coefficients for T Ranging from Minimum to Maximum

Contrary to the results on absolute bias, the Mean Squared Error for the BMLP BTSCS model is larger than the one of the pooled probit, Beck et al.'s (1998) standard BTSCS approach and the two frequentist multi–level models. However, the BMLP BTSCS model's MSE decreases with both an increase in N and T: while the MSE is largest for $N = 10$ and $T = 10$ (≈ 3.16) compared with all other models, it is equivalent or even smaller for $N = 50$ and $T = 50$ (≈ 0.05).

Standard Errors

The problem of contemporaneous correlation in BTSCS data can lead to incorrect standard errors and therefore affect hypothesis testing since the assumption of independent observations in non–linear models is violated (Beck et al. 1998, 1261–1264; Beck 2001a, 128–129; Shor et al. 2007, 165–172). Although we prefer standard errors to be as small as possible since this reflects a higher confidence of our estimates, smaller standard errors can be incorrect if there is in fact a high level of variability in the sample (Carsey and Harden 2013, 99–102). Despite being a concept of frequentist statistics, Bayesian "standard errors" as a measure of estimate uncertainty can be derived from the standard deviation of the posterior distribution (Lee et al. 2017, 868). In order to ascertain whether the BMLP BTSCS model yields correct standard errors, I follow Carsey and Harden (2013) by applying the so called "standard deviation" method, which makes it possible to assess whether the standard errors of a model's parameter estimate reflect this parameter's variability across different samples (Carsey and Harden 2013, 99–102).

Ideally, the mean of the standard errors calculated in each respective Monte Carlo trial for each model's β–coefficient estimate should be close to the standard deviation of the distribution of the simulated 250 coefficient estimates (Carsey and Harden 2013, 99–102). This similarity can be assessed by comparing these two quantities (Carsey and Harden 2013, 99–102). In order to facilitate the presentation of the results, I simply subtracted the models' average standard errors for each value of N and T from the observed standard deviation of the coefficient estimates of all Monte Carlo trials such that a value equal to or close to 0 indicates correct standard errors. Figure 16 and 17 show the variation of this difference over both N and T[26].

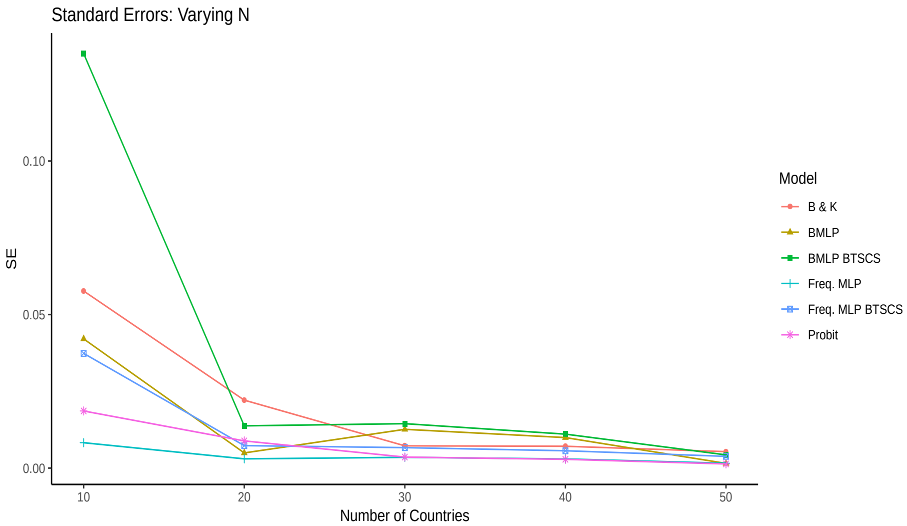


Figure 16: Average Difference Between the SEs of the β–Coefficients Compared with the Standard Deviation of the Sampling Distribution for N Ranging from Minimum to Maximum

[26]Exact numbers are presented in Table 5 in the Appendix. Also this table was created with the `R`–packages `dplyr`; `reshape2`; `splitstackshape` and `xtable` (Dahl 2016; Mahto 2018; Wickham 2007; Wickham et al. 2017).

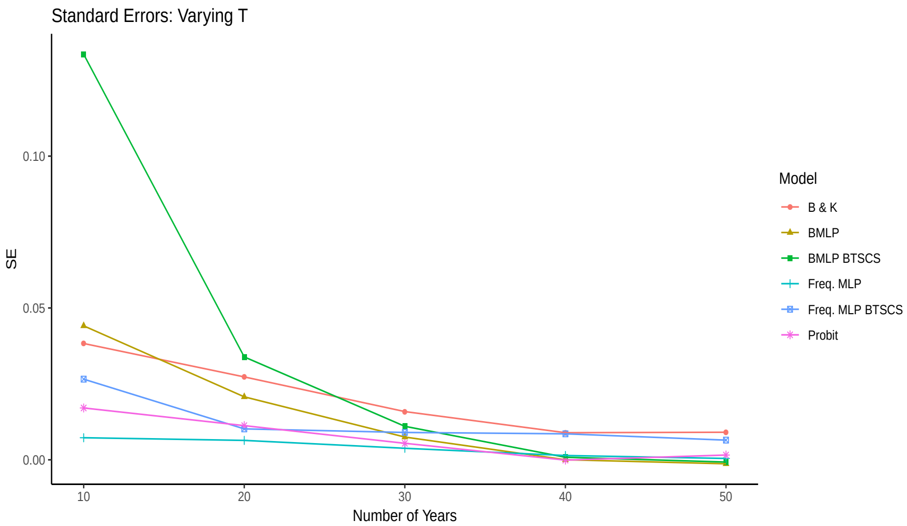


Figure 17: Average Difference Between the SEs of the β–Coefficients Compared with the Standard Deviation of the Sampling Distribution for T Ranging from Minimum to Maximum

The small values on the y–axis of the results presented in Figure 14 and 15[27] reveal that the coefficient standard errors of *all* models come close to the true coefficient variability. This is surprising since pooled probit models are usually known to underestimate coefficient variability for BTSCS data (Beck et al. 1998, 1263). Furthermore, the coefficients of the BMLP BTSCS model yields larger standard errors than the other models across all values of N which means that the estimated variability is larger than the true sample variability (Shor et al. 2007, 175). However, as Table 8 (see Appendix) shows, this "underconfidence" diminishes with increasing N and T, even for a small number of countries, i.e. $N < 30$.

Finally, in order to provide a more intuitive comparison between the various models, I follow Stegmüller (2013) and present a graphical illustration. Since it is not possible to present 25 plots for every combination of N and T ranging from 10 to 50[28], respectively, I focus on the results for $N = 20$ and $T = 20$, $N = 30$ and $T = 20$ and $N = 30$ and $T = 30$ to evaluate the BMLP BTSCS model's results compared with all other models' results for N below and equal to the critical value of 30 (Maas and Hox 2004, 127–130; Stegmüller 2013, 748–751). The reason for the choice of these value combinations is that they resemble the structure of my dataset and two EU research applications: First, the

[27]For the creation of the graphs, the `R`–packages `dplyr`, `gdata` and `ggplot2` were used (Warnes et al. 2017; Wickham 2016; Wickham et al. 2017).

[28]For the remaining figures, see Figures 38 to 42 in the Appendix.

Eurozone comprises nineteen countries and exists since nineteen years, which both matches approximately $N = 20$ and $T = 20$ (European Union 2018b). Second, the datasets used for this analysis comprises twenty–five countries and sixteen years, which is similar to $N = 30$ and $T = 20$. Third, the EU as a whole currently has twenty–eight members and existed for twenty–five years, which comes close to the case of $N = 30$ and $T = 30$ (European Union 2018a).

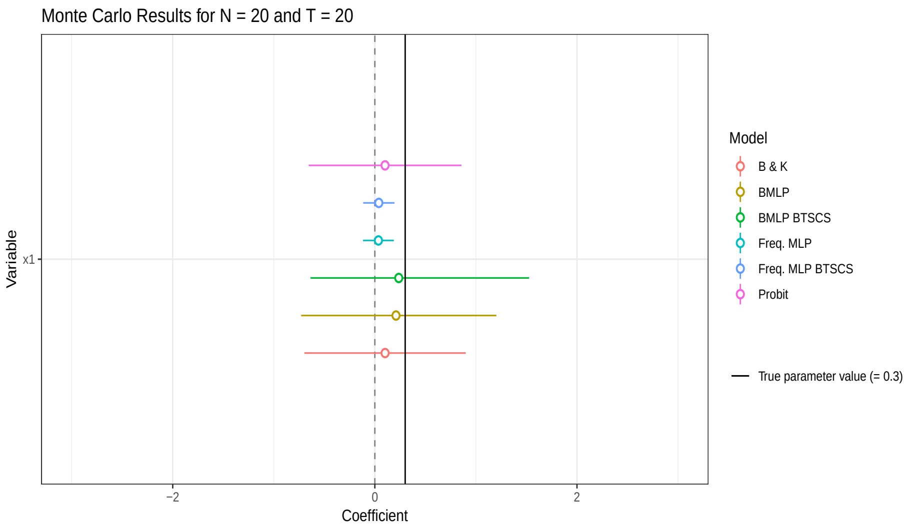


Figure 18: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 20$ and $T = 20$

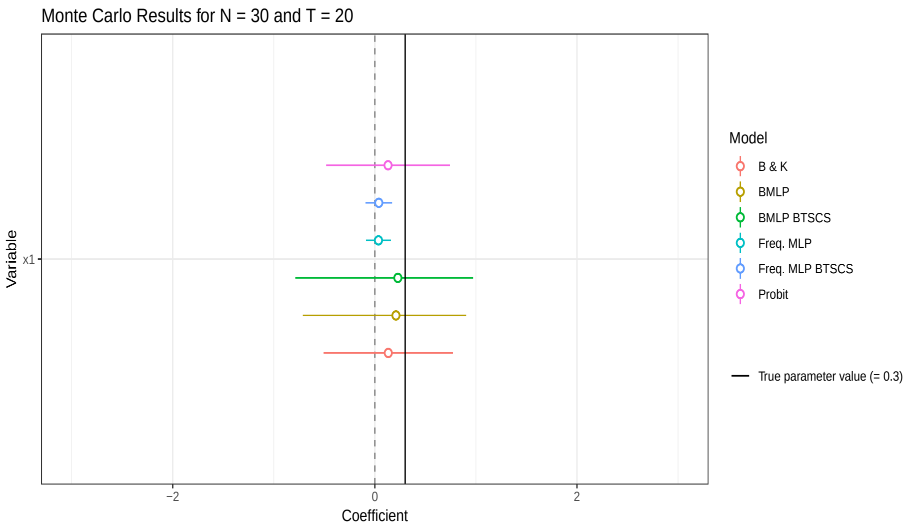


Figure 19: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 30$ and $T = 20$

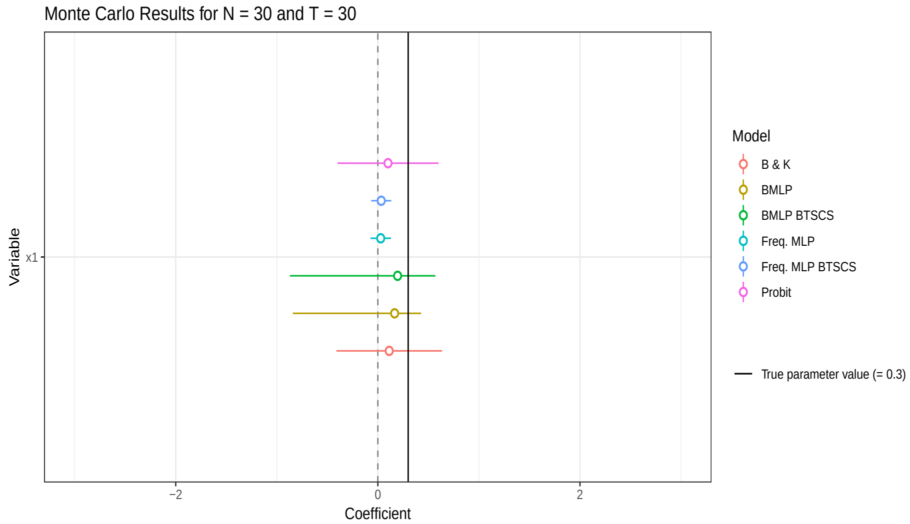


Figure 20: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 30$ and $T = 30$

Figure 18 to 20[29]plot both point estimates and 95 % confidence/Bayesian credible intervals of the β parameters of all models. The solid vertical line indicates the true parameter value of β, 0.3, to make deviations from this value more clear. For $N = 20$ and $T = 20$ – namely a small sample size including panel heteroskedasticity and contemporaneous correlation – the BMLP BTSCS model's coefficient is closest to the true value compared with the coefficients of the other models. Particularly the coefficients of the pooled probit model and the two frequentist multi–level models deviate from the true parameter value. Similar to the findings of Pang (2010) and Stegmüller (2013), the BMLP BTSCS model's credible intervals are much larger than those of the confidence/credible intervals of the other models, which makes it less likely to reject a true null hypothesis (Pang 2010, 486–489; Shor et al. 2007, 176; Stegmüller 2013, 756–758). The same is true for the results where $N = 30$ and T is either equal to 20 or 30: Figure 17 and 18 show that the point estimate of the BMLP BTSCS model is closest to the true coefficient value and credible intervals are not overly confident. This suggest that the BMLP BTSCS model is an appropriate choice for BTSCS data, particularly for samples with a small N such as Eurozone/EU analyses.

Note that the results of both the frequentist multi–level model (country–level random intercept only) and the frequentist Model with the BTSCS specification are contrary to the expectation: Despite accounting for panel heteroskedasticity, both models yield biased

[29]The figures were created using `R`'s `ggplot2` and `tidyr` (Wickham 2016; Wickham and Henry 2018).

estimates, while the frequentist multi–level BTSCS model's confidence intervals are much smaller than expected. This is most likely due to convergence problems that persisted during the experiment despite scaling the independent variable (Long 1997, 59–60). Since the dependent variable is completely random and is therefore neither censored nor one of its categories has very few cases, nonconvergence is most likely due to small sample size (Long 1997, 59–60). This once again confirms the problematic application of ML estimators on small N data.

It is important to highlight here that the experiment has been conducted using balanced BTSCS data, while many BTSCS datasets – including the datasets used for this analysis – comprises unbalanced BTSCS data. However, the BGLM is still applicable on unbalanced data with even a very small number of observations per unit (1 or 2) as these data still contain information that improve the estimation of coefficients and variance parameters (Gelman and Hill 2006, 276).

In sum, the Monte Carlo experiment has confirmed that the BMLP BTSCS model provides unbiased estimates and correct standard errors even for small samples where $10 < N/T \leq 30$, which makes it applicable for the analysis of the EU BTSCS data used for the investigating the effect of public support and transparency on sanctioning SGP violations.

8 Analysis: The Effect of Public Support and Transparency on EDP Initiations

After the confirmation of the suitability of the BGLM to be applied for this investigation, I will present the statistical analysis in this section. However, before discussing the results of this investigation, I will present some descriptive statistics of the dependent and key independent variables [1].

8.1 Descriptive Statistics

Figure 21 shows the number of EDPs (measured as transferred opinions on the existence of an excessive deficit) since 2002 in the EU per country. In the sample of "fiscal sinners", there are only three countries that have never experienced an EDP, namely Denmark, Finland and Hungary. This result is particularly surprising for Hungary since it is has already violated the pact eight times, as shown in Figure 25. This once again indicates that many violations remain unsanctioned by the Commission. All other countries experienced at least one EDP. Malta is a clear exception in this respect since it already experienced three EDPs. However, for Malta only a few of its six committed violations were sanctioned. Overall, this figure illustrates that although the SGP has only been in force for a very

[1] All graphs in this section were created using the `R`–packages `car`, `ggplot2` and `scales` (Fox and Weisberg 2011; Wickham 2016; Wickham 2017). Table 9 to 21 in the Appendix present summary statistics of all variables included in the dataset created for this analysis. For simplicity, I will focus on the dependent and key independent variables in this section. The tables were created with the `R`–packages `dplyr` and `xtable` (Dahl 2016); Wickham et al. 2017).

short time, the majority of EU–members got already sanctioned for violating it, which is unsurprising given the increasing debt and deficit levels in the EU (see Figures 3 and 4). Therefore, there variation across countries is not very strong for this variable.

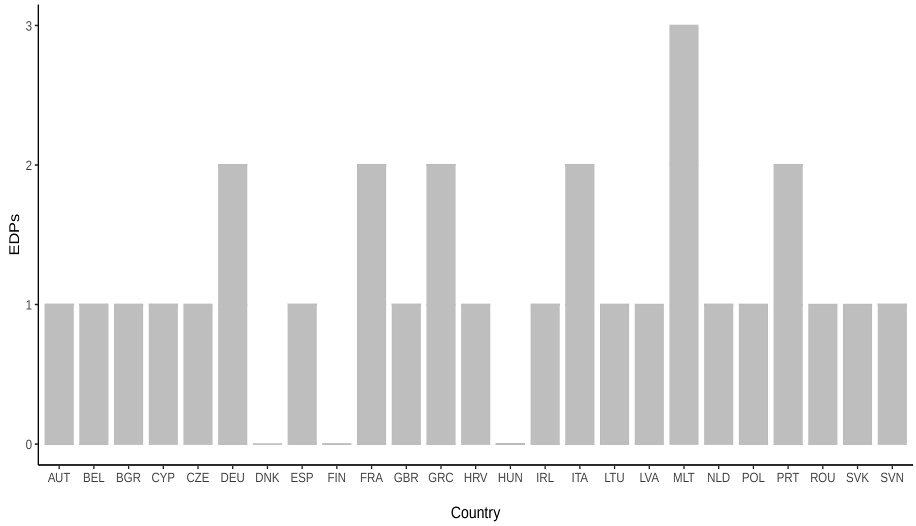


Figure 21: EDPs (Opinion) per Country

Figure 22 shows the number of EDPs using the alternative specification of the dependent variable, the publishing of the Commission report per country and year. For this measure, Italy is besides Malta the country experiencing the most EDP initiations. Furthermore, Finland is no longer among the countries that experienced no EDP for this measure. However, this graph also confirms that most EU countries have already experienced at least one EDP.

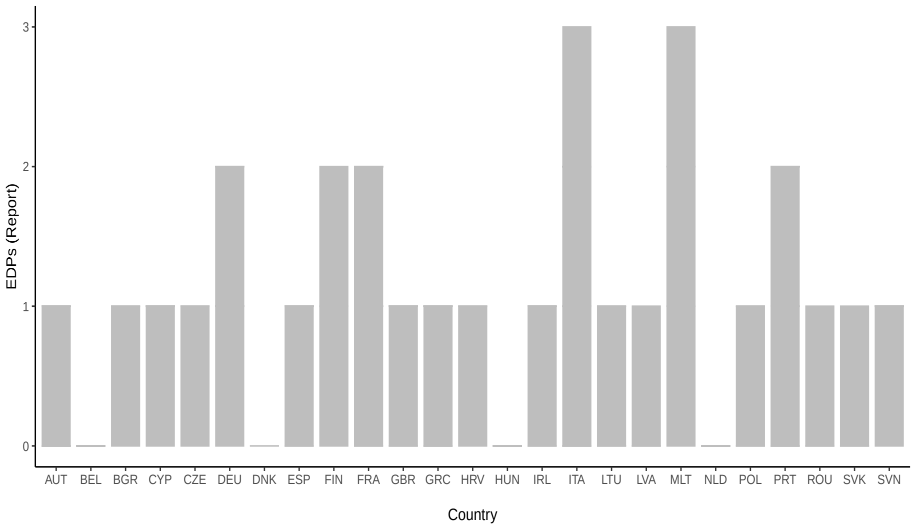


Figure 22: EDPs (Reports) per Country

Figure 23 illustrates the number of EDPs since 2002 per year [2]. This figure is quite different from the one before, since there is much more variation. One can see that the vast majority of EDPs was initiated in 2009, followed by 2004. This is also true for the alternative specification of the dependent variable plotted in Figure 24.

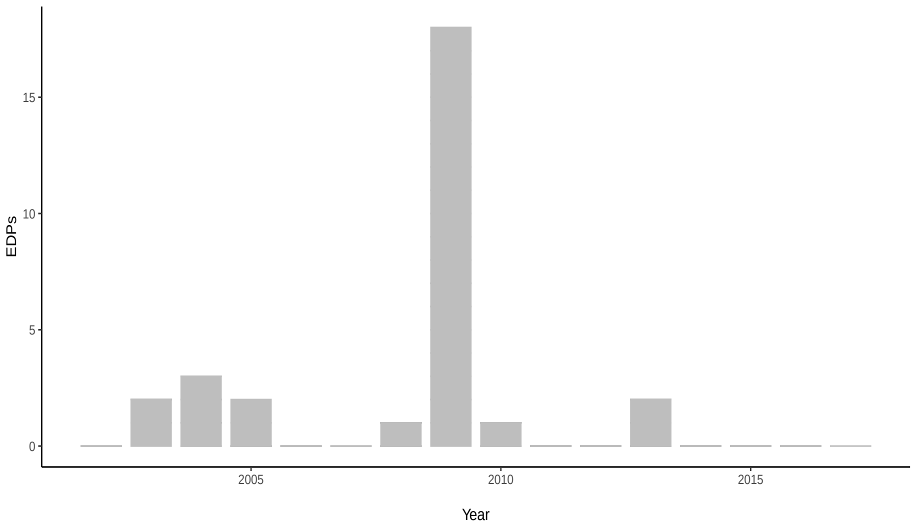


Figure 23: EDPs (Opinion) per Year

[2]For a tabular overview of the number of EDPs per country, see Table 6 in the Appendix.

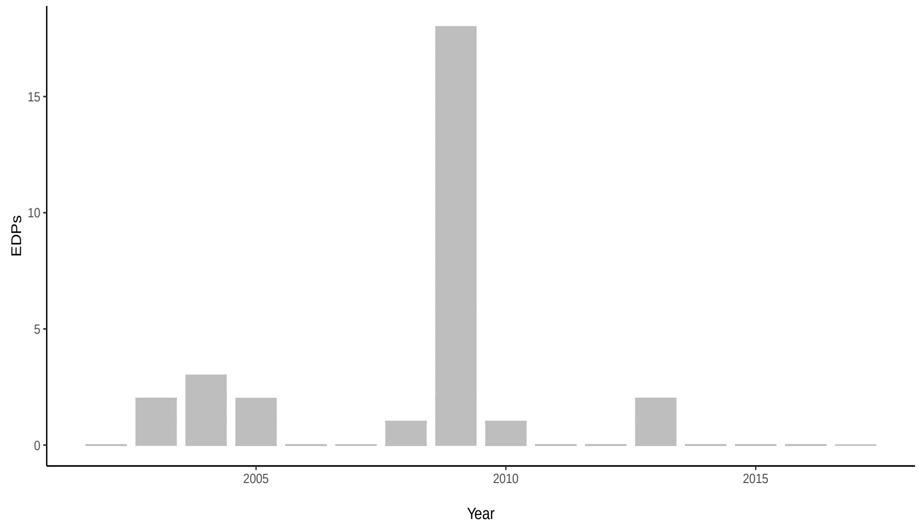


Figure 24: EDPs (Reports) per Year

When comparing this result with the number of SGP violations per year shown in Figure 26 reveals that 2009 was indeed the year were most SGP violations were committed. However, many SGP violations also took place in the remaining years, while much fewer EDPs were opened by the Commission.

Since many European states were hit by the global financial crisis in 2009, the observations in Figure 26 contradict hypothesis 1 because these economic turbulences did not stop the Commission from initiating an EDP against several member states (Lane 2012, 55–56). Figure 23 and Figure 24 also demonstrate that the given sample of "fiscal sinners" contains much more variation on the dependent variable over time than across countries, which makes it even more important to model this variation by an additional year–specific random intercept in the multi-level model.

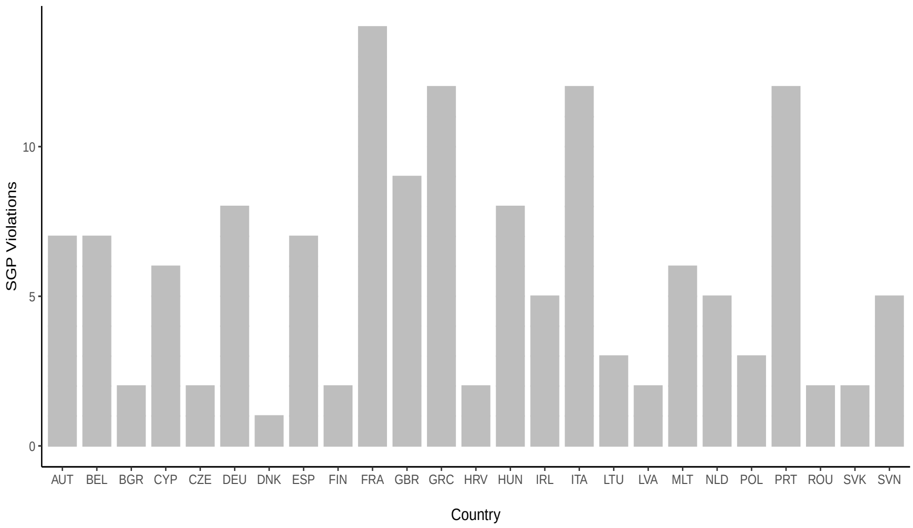


Figure 25: Number of SGP Violations per Country

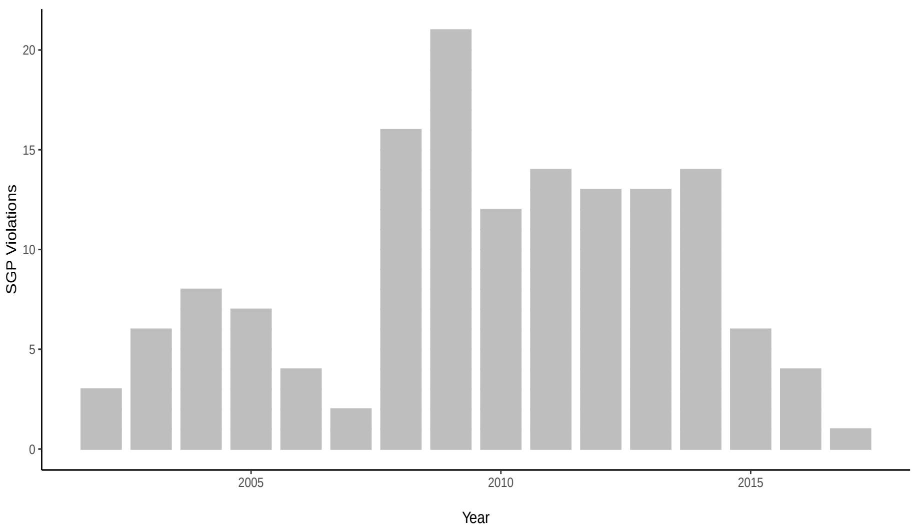


Figure 26: Number of SGP Violations per Year

In sum, as already illustrated in Table 1 , breaking the SGP does not necessarily seem to lead to an opening of an EDP by the European Commission. Overall, the descriptive statistics shown here confirm the suspicion that the Commission's sanctions are not exclusively based on the violation of the SGP's fiscal rules.

The key argument of this investigation is that the Commission is driven by political considerations such as public support and transparency about a country's economic situation when sanctioning SGP violations. As can be seen in Figures 27 to 31, these factors

Figure 27: Trust in European Commission per Country and over Time

vary considerably across countries and over time. First of all, the absolute support for the Commission measured by the share of respondents who express trust in the European Commission shown in Figure reffig:comtrust decreases in all countries for most of the years, even though there is a sligh increase before 2015. Over time, Belgium is the country in which most citizens trust the Commission, while the level of trust is – unsurprisingly – the lowest in Greece. At the same time, Belgium experienced a lower number of EDPs (one EDP while violating the SGP seven times) than Greece (two EDPs out of twelve violations), which matches with the argument that higher levels of support for the Commission lead to more EDP openings. However, the difference in the number of EDPs is, admittedly, not very large between these two high– vs. low–support countries.

Furthermore, the relative support for the Commission compared with national governments presented in Figure 28 shows a slight negative trend over time, which indicates that support for the Commission and consequently the SGP seems to slightly diminish compared with governmental overspending. The positive values on the y–axis indicate that support for the Commission (and therefore the SGP) exceeds support for governmental overspending for most countries in the dataset. However, overall there is no country in which the support for the Commission is consistently much higher than support for the national government throughout the time period to be investigated. However, in Malta, the country which experienced the most EDPs, support for the Commission consistently (but diminishing) exceeds support for the national government. In this sense, Malta matches the theoretical expectation that high trust results in more EDP initiations. By contrast, the

Figure 28: Relative Public Support per Country and over Time

difference between support for the Commission and support for the national government is not very large for Austrian voters. They even seem to support their national government more than the Commission around 2009/2010. At the same time, out of Austria's seven SGP violations, only one was sanctioned. This pattern confirms the theoretical expectation that lower (relative) support for the Commission makes the Commission more hesitant to sanction SGP violations.

Figure 29: Share of Respondents with a Positive Image of the EU per Country and over Time

Finally, as previously mentioned, it is possible that voters might not support the Commission, but the EU and its rules including the SGP in general. Figure 29 shows that despite a negative trend in the share of respondents who expressed having a positive image of the EU, this share increased around 2013 to 2015. In Romania and Bulgaria, most respondents have a very good impression of the EU, while Greece and the UK are among the states with the lowest share of respondents who perceive the EU as positive.

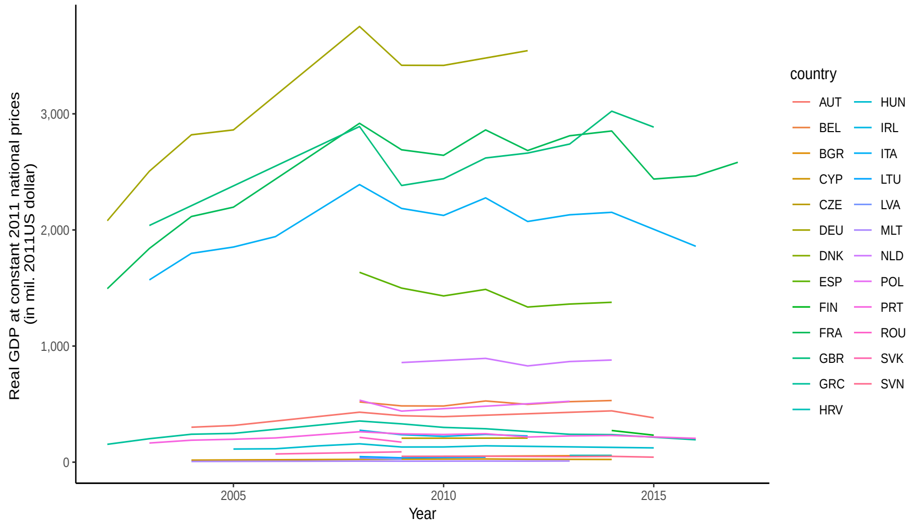


Figure 30: Real GDP of EU States per Country and over Time

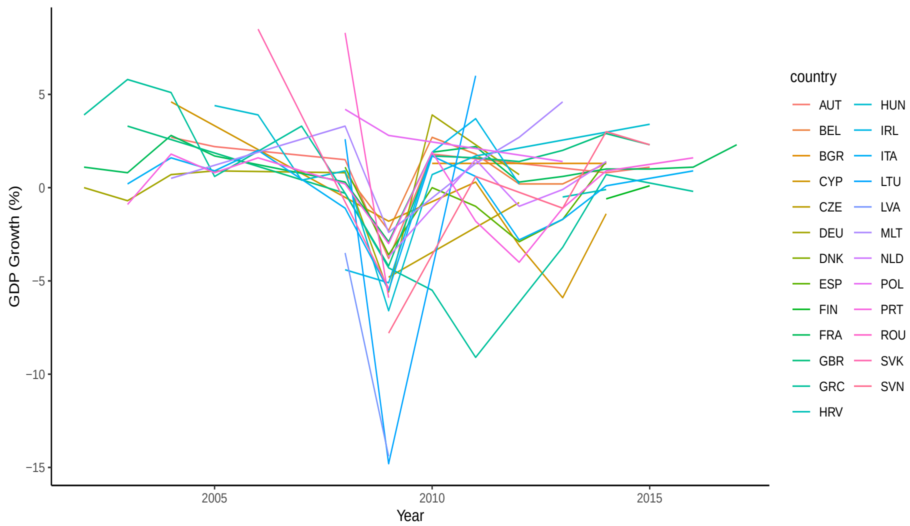


Figure 31: GDP Growth of EU States per Country and over Time

While there is much variance in all indicators measuring public support for the Commission between countries and over years, Figure 30 shows much less variation in transparency of a country's economic situation measured by real GDP. The graph depicts a remarkably stable economic situation for all EU member states despite the experiences of the recent European Sovereign Debt crisis. Here, Germany exhibits by far the highest real GDP over the years, while Malta maintains the lowest one. In addition, no country seems to experience sudden and large declines in real GDP. The absence of such unforeseen economic downturns suggests that transparency for the Commission – at least according to the criteria of the SGP – is given in most member states. This picture changes when considering the development of the second transparency measure, GDP Growth, in the selected EU member states over time, which is depicted in Figure 31. Here, we can see a strong decline for all member states in the crisis year 2009, which is sharpest for Latvia and Lithuania. As a result, while transparency about deliberate overspending is given when considering real GDP, this picture is somewhat less clear when considering economic growth. This is once more not in line with hypothesis 1 as most EDPs were initiated when several countries experienced a recession.

Upon first glance, the descriptive statistics seem to not support hypothesis 1. The presentation of the results of the statistical analysis in the following section will clarify whether the postulated relation between public support, transparency and EDP initiations holds.

8.2 Regression Results

	Dependent Variable											
	EDP – Opinion						*EDP – Report*					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Trust in European Commission × Real GDP	–0.00 (0.00)						–0.00 (0.00)					
Relative Public Support× Real GDP		–0.00 (0.00)							–0.01 (0.01)			
Trust in European Commission × GDP Growth			–0.01 (0.74)					0.43 (0.60)				

Table 4: Overview Regression Results

	Dependent Variable											
	EDP – Opinion							*EDP – Report*				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Relative Public Support× GDP Growth				–0.74 (0.58)						–0.86 (0.58)		
Positive Image of the EU × Real GDP					0.00 (0.00)						–0.00 (0.00)	
Positive Image of the EU × GDP Growth						–0.46 (0.55)						–0.12 (0.47)
Trust in European Commission	18.17 (9.83)		11.29 (6.03)				14.58 (7.40)	8.05 (3.51)				

Table 4: Overview Regression Results

	Dependent Variable											
	EDP – Opinion								*EDP – Report*			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Relative Public Support		–1.73 (3.19)		–2.25 (2.89)					–4.75 (4.45)	–6.46 (4.14)		
Positive Image of the EU					7.28 (3.88)	8.63 (3.56)					5.17 (3.52)	6.01 (2.51)
Real GDP	0.00 (0.00)	–0.00 (0.00)			–0.00 (0.00)		0.00 (0.00)		–0.00 (0.00)		0.00 (0.00)	
GDP Growth			–0.13 (0.37)	0.09 (0.14)		0.05 (0.25)		–0.37 (0.30)		0.05 (0.15)		–0.12 (0.22)
Powerful Memberstate	–0.37 (2.19)	1.21 (1.43)	1.05 (1.24)	0.51 (1.03)	–0.35 (1.40)	0.45 (0.87)	0.34 (1.89)	0.25 (0.95)	1.89 (2.56)	–0.08 (1.31)	–0.18 (1.39)	–0.15 (0.83)
Non–Compliance	0.37 (0.30)	0.35 (0.30)	0.28 (0.25)	0.34 (0.29)	0.29 (0.24)	0.24 (0.23)	0.06 (0.12)	0.01 (0.09)	0.13 (0.20)	0.05 (0.13)	0.05 (0.09)	–0.00 (0.08)
Eurozone Member	–2.41 (1.42)	–1.41 (0.92)	–2.10 (1.32)	–1.37 (0.98)	–1.50 (0.86)	–1.83 (0.98)	–1.47 (1.02)	–1.22 (0.81)	–2.02 (1.50)	–1.52 (1.17)	–1.10 (0.81)	–1.07 (0.73)

Table 4: Overview Regression Results

	Dependent Variable											
	EDP – Opinion								*EDP – Report*			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Election	0.19	0.41	0.34	0.43	0.29	0.51	–0.04	–0.01	0.11	0.11	0.07	0.12
	(0.67)	(0.53)	(0.65)	(0.56)	(0.57)	(0.60)	(0.51)	(0.47)	(0.56)	(0.50)	(0.44)	(0.44)
Intercept	–13.47	–6.43	–10.08	–5.85	–7.10	–7.35	–7.65	–3.66	–1.53	1.01	–2.34	–1.95
	(6.99)	(4.79)	(4.94)	(4.50)	(4.31)	(3.95)	(4.62)	(2.43)	(3.89)	(2.74)	(2.56)	(1.83)
σ_α	3.68	1.52	2.84	2.16	1.15	1.23	4.15	2.22	9.12	5.02	2.21	1.51
	(4.07)	(1.71)	(3.70)	(2.49)	(1.42)	(1.46)	(4.02)	(2.00)	(10.57)	(5.08)	(2.22)	(1.45)
σ_γ	15.75	16.72	12.05	16.38	10.36	9.75	4.59	2.56	12.09	5.53	2.88	1.61
	(17.87)	(18.33)	(15.70)	(17.73)	(14.44)	(13.57)	(5.68)	(3.20)	(14.99)	(6.20)	(3.57)	(1.85)

Table 4: Overview Regression Results

Table 4 present the results (coefficient and standard deviation) of all regression models including robustness checks. Column 1 and 2 include the coefficient estimates for the interactions between the two public support indicators *Trust in European Commission* and *Relative Public Support* and real GDP, respectively[3].

[3]Due to the complexity of the BMLP BTSCS model, I ran all models with three chains iterating 800,000 times with a burn–in of 400,000 and a thinning interval of 400, which leaves 1,000 simulation draws saved for each chain (Gelman and Hill 2006, 418; Gelman et al. 2013, 282–283; Shor et al. 2007, 173). Here, values close to 1 of the Gelman–Rubin $\hat{R}$ provided by the `coda`—package confirmed chain convergence (Cowles and Carlin 1996, 884–885; Gelman et al. 2013, 284–285; Plummer et al. 2006; Shor et al. 2007, 173). Besides checking for chain convergence, I follow Gelman and Hill (2006) and Gelman et al. (2013) in conducting posterior predictive checking (Gelman and Hill 2006, 513–521; Gelman et al. 2013, 143–150). For this purpose, I simulated replicated datasets $p(y_{rep}|X, \beta = \beta^s)$, the so called posterior predictive distribution, for the $s = 3,000$ simulated parameter draws β^s given the matrix of independent variables X (Gelman and Hill 2006, 513). From the posterior predictive distribution, I generated two discrepancy variables $T(y_{rep}, \beta)$, mean and standard deviation, and compare them with the mean and standard deviation of the observed data $T(y, \beta)$ (Gelman and Hill 2006, 513–521). Furthermore, I computed Bayesian p–values defined as $p = Pr(T(y_{rep}, \beta) > T(y, \beta|y)$ that indicate which proportion of the 3,000 simulations for which the test quantity is more extreme than the actual data (Gelman and Hill 2006, 514); Gelman et al. 2013, 146). Figures 73 to 96 show histograms of the mean and standard deviation for $T(y^{rep})$ for each model. The vertical red lines indicate the value of the observed mean and standard deviation of y, respectively, which corresponds to mean and standard deviation of the dependent variable EDP (Gelman and Hill 2006, 520–521). All graphs indicate a fairly good match between the mean and standard deviation of the replicated data of all three models and the actual mean and standard deviation of the dependent variable (Gelman and Hill 2006, 520–521). Moreover, the value of the Bayesian p–values (see Table 22 in the Appendix) indicate a close similarity between the replicated and the actual data since their values are close to 0.5 for all models (Ntzoufras 2009, 342. Finally, Table 23 in the Appendix reports the Deviance Information Criterion (DIC) of each model which indicates each model's out–of–sample predictive error (Gelman and Hill 2006, 525). Generally, lower values of the DIC indicate a better model fit, whereby Model 6 is the best model for out–of–sample predictions as it has the lowest DIC (Gelman and Hill 2006, 526).

The coefficient sign is equal to zero for both the interaction *Trust in European Commission* $\times$ *Real GDP* and the interaction consisting of *Relative Public Support* and *Real GDP*, which indicates that the mean of the posterior distribution obtained from the Bayesian model is close to zero (Western and Jackman 1994, 418). Since the 2.5% and 97.5% percentile of the posterior distribution equal –0.01 and 0.01 for the coefficients for both interaction terms, there is a 95% probability that the parameter value lies between these values (Martin 2008, 503; Western and Jackman 1994, 418).

However, it should be stressed that it is not possible to infer whether public support and transparency have a meaningful effect on the initiation of an EDP from the magnitude of the interaction terms' coefficients (Brambor et al. 2006, 74). This is because the coefficient of interaction terms in non–linear models is not equal to the marginal effect of the interaction term (Ai and Norton 2003, 122–129; Brambor et al. 2006, 77)[4]. Moreover, as the interaction effect for an interaction term consisting of two continuous variables x and z can be described by the cross derivative of the expected value of the dependent variable

$$E[Y] = \Phi(\beta_1 x + \beta_2 z + \beta_3 xz + X\beta) = \Phi''(\cdot) \tag{8.1}$$

,

$$\frac{\partial^2 \Phi(\cdot)}{\partial x \partial z} = \beta_3 \Phi'(\cdot) + (\beta_1 + \beta_3 z)(\beta_2 + \beta_3 x)\Phi''(\cdot) \tag{8.2}$$

,

which remains non–zero and equal to

$$\frac{\partial^2 \Phi(\cdot)}{\partial x \partial z} = \beta_1 \beta_2 \Phi''(\cdot) \tag{8.3}$$

,

[4]Note that it is also not possible to interpret the coefficients of the constituent terms of the interaction terms included in all models as the average effect on the dependent variable (Brambor et al. 2006, 71–72). This is because they only capture the effect of each public support indicator on the opening of an EDP when real GDP is equal to zero (and vice versa) (Brambor et al. 2006, 71–72). The only way to measure the average effect of the constituent terms on the dependent variable is to run an unconditional model without an interaction term that consists of these constituent terms (Brambor et al. 2006, 72). Since only conditional hypotheses are of interest for this analysis, I did not conduct such an additional analysis.

even if the coefficient for the interaction term, β_3, is equal to zero (Ai and Norton 2003, 124).

Since the table results convey little information for hypothesis testing, I follow Brambor et al. (2006), Gelman and Hill (2007) and King et al. (2000) and present a graphical illustration of the results (Brambor et al. 2006, 74; Gelman and Hill 2006, 94; King et al. 2000, 347–351). In line with King et al. (2000), I simulate expected values (see equation 8.1), which not only allows me to assess the average effect of each public support indicator on EDP initiation for different levels of transparency, but also to determine the uncertainty around this effect (King et al. 2000, 347–351). For frequentist models, this simulation procedure usually involves drawing randomly M β parameters from a multivariate normal distribution that takes the parameters' point estimates and variance–covariance matrix as mean and standard deviation (King et al. 2000, 349–350). For a Bayesian model like the one used here, simulating expected values is much more convenient since it is possible to directly gather the M parameter estimates from the posterior distribution, which already comprises $M = 3,000$ simulation draws in total[5]. (Gelman and Hill 2006, 418; Gelman et al. 2013, 282–283; King et al. 2000, 349–352). These parameter draws, in turn, can vary in their size, thereby reflecting the *estimation* uncertainty about their exact value which arises from having only a finite number of observations at hand (King et al. 2000, 349). Besides estimation uncertainty, there also exists *fundamental* uncertainty, which stems from the fact that there might be always a factor not considered in the set of independent variables of the model that affects the outcome variable (King et al. 2000, 349). The large number of simulations ($M = 3,000$) ensures that the expected values average over fundamental uncertainty, while estimation uncertainty remains left (King et al. 2000, 350–351).

Besides the random parameter draws, it is also necessary to choose meaningful values for each explanatory variable included in the model for the calculation of the expected values (King et al. 2000, 351). For this purpose, I set up two "scenarios", where each public support measure ranges from minimum to maximum, while the conditioning variable transparency is set equal to either its maximum (indicating high transparency for the Commission) or its minimum (indicating low transparency for the Commission). For the control variables, I fixed the value for the count control variable Non–Compliance equal to its mean, while the remaining binary control variables were held constant at value 1

[5]This number stems from running each model with three Markov Chains of 1,000 iterations, respectively (Gelman and Hill 2006, 418; Gelman et al. 2013, 282–283; King et al. 2000, 349–352).

(King et al. 2000, 355). Consequently, the quantity of interest to be simulated here is the probability of an EDP initiation if public support changes from minimum to maximum and transparency is high, given a powerful Eurozone memberstate with a mean level of non–compliance and an election taking place. Finally, the expected values of this quantity are obtained via multiplying the coefficient draws with the predefined covariate values and plugging them into a standard normal cumulative distribution link function, which is the systematic component of the probit model (King et al. 2000, 348–355; Wooldridge 2002, 457–458; Wooldridge 2009, 575–576). The expected values for the two scenarios for the first model in which absolute public support is measured by the amount of trust in the Commission and transparency by real GDP are shown in Figure 32 and 33, respectively.

The solid line in Figure 32 shows the probability of opening of an EDP for high levels of transparency (namely, high levels of real GDP) when trust in the European Commission ranges from minimum to maximum. As one can see, higher values of public support seem to have a meaningful effect on the probability of EDP openings if transparency is high.

However, the vertical lines in the graph show the estimation uncertainty of these expected values (King et al. 2000, 355). The probability of an EDP initiation ranges from 0 to 1 for all values of the variable *Trust in European Commission* and high levels of transparency (King et al. 2000, 355). This result implies that the launch of an EDP by the European Commission is very uncertain even if both transparency and public support are high. The reason for this high estimation uncertainty stems from the low number of observations ($N = 406$) used for this investigation (King et al. 2000, 355).

Note that the small N problem persists also in a Bayesian model: for a large N that moves towards infinity, the posterior collapses on a spike over the true parameter value because the information in the data overwhelm the prior (Jackman 2009, 30–31). The combination of uninformative priors for the coefficients of the independent variables which state a diffuse effect of these variables on EDP and a small dataset leaves a larger variance of each parameter's spike because the information in the data are not sufficient to overrun the uncertainty induced by the prior (Jackman 2009, 16). This is reflected as estimation uncertainty in the model.

Figure 33 illustrates the counter scenario, namely the expected probability of opening an EDP if transparency is low (that is, low levels of real GDP) while the values of the public support–indicator increase. Interestingly, the horizontal line in this graph implies that public support for the Commission seems to still have a positive effect on EDP initiations, even if transparency is low. This supports the first theoretical argument that public support is crucial for NMI sanctioning, but contradicts the second theoretical argument that this effect would depend on the level of transparency. Moreover, from the vertical lines one can see that the estimation uncertainty of an EDP initiation is in this graph much higher for high levels of trust in the European Commission compared with lower levels of trust.

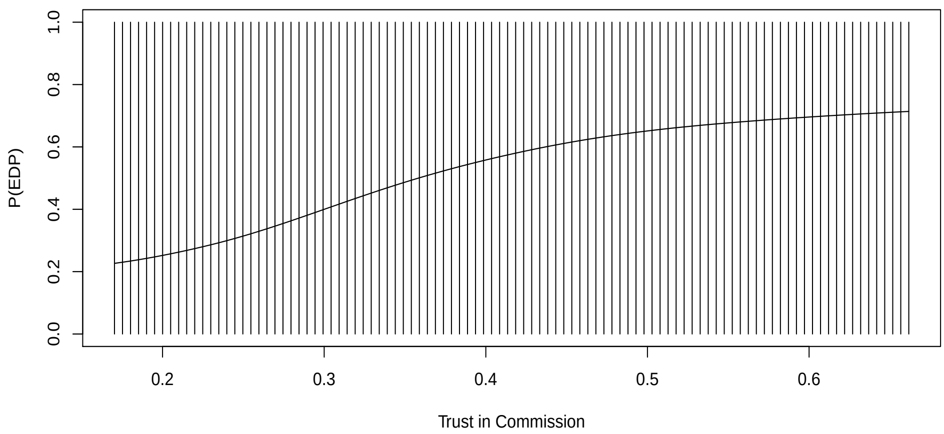


Figure 32: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *Real GDP* on EDP (Opinion) Initiation (*Real GDP* = high)

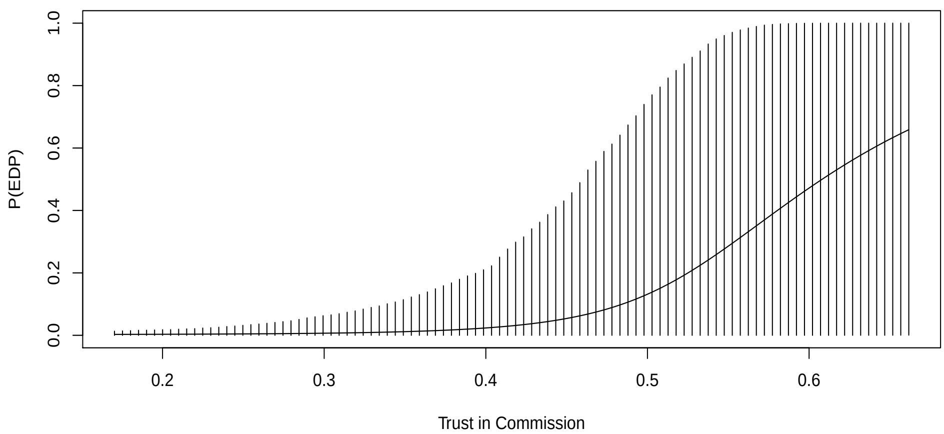


Figure 33: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *Real GDP* on EDP (Opinion) Initiation (*Real GDP* = low)

The results of the first model indicate that higher values of absolute support for the Commission increase the probability of an EDP initiation regardless the level of transparency, albeit with a high estimation uncertainty. Figure 34 shows the effect of *relative* amount of public support for the Commission – measured by the differential between the variables *Trust in European Commission* and *Trust in National Government* – on EDP initiations given high transparency. From this graph, we can also observe that if trust in the Commission exceeds the level of trust for the national government, the probability for an EDP initiation slightly increases given a high level of transparency. Moreover, Figure 34 seems to support Köhler et al.'s (2018) predictions that only if voters' support for compliance with the SGP is higher than their support for governmental overspending, the Commission is more willing to sanction violations of the SGP because its enforcement costs are low in this case. Nevertheless, here, the probability of an EDP initiation ranges between 0 and 1 for most values of the variable *Relative Public Support* given that transparency is high, thereby indicating high estimation uncertainty of the results due to the small dataset.

As for the first Model, Figure 35 depicts how the effect of *Relative Public Support* affects EDP initiations if transparency is low, while the uncertainty of an EDP initiation is lowest for low levels of *Relative Public Support*. Somewhat counter–intuitively, an increase in the second public support indicator seems to reduce the probability of an EDP initiation if no sufficient level of transparency is given, and the uncertainty of the probability of sanctioning also decreases here. In contrast to the first model, the effect of *Relative Public Support* is clearly dependent on the amount of transparency, given that the effect of public support on EDP initiations completely changes for low levels of transparency. The conclusion drawn from this model is therefore that even if voters support the Commission and the SGP more than governmental overspending, the Commission will refrain from opening an EDP as it cannot be certain that violating the SGP's rules was committed deliberately. The results of this model indicate that the positive effect of rising levels of public support depends on high levels of transparency. Unfortunately, the data still do not provide a sufficient amount of information to reduce estimation uncertainty of this result.

Regarding the BTSCS specification, one can see from Table 4 that the variability among years is larger than the variability among countries in both models (Gelman and Hill 2006, 458). This is further confirmed by the caterpillar plots[6] of the country–level random effects

[6] All caterpillar plots in this section and the appendix have been created with the

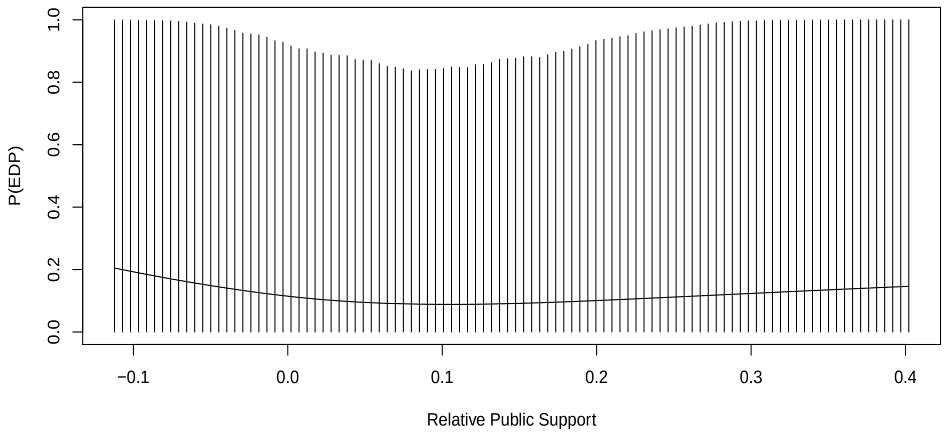


Figure 34: Expected Values for the Effect of the Interaction between *Relative Public Support* and *Real GDP* on EDP (Opinion) Initiations (*Real GDP* = high)

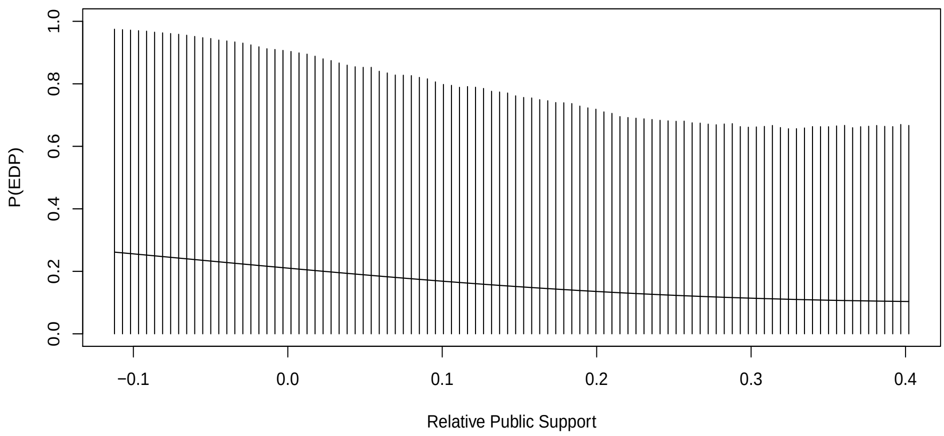
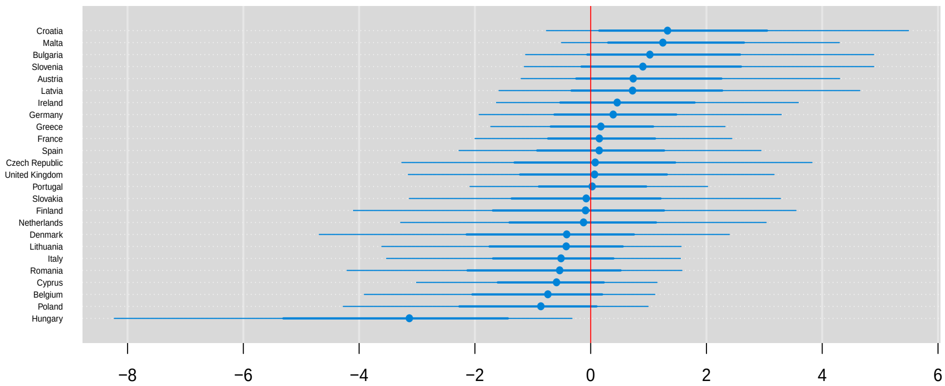


Figure 35: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *Real GDP* on EDP (Opinion) Initiations (*Real GDP* = low)

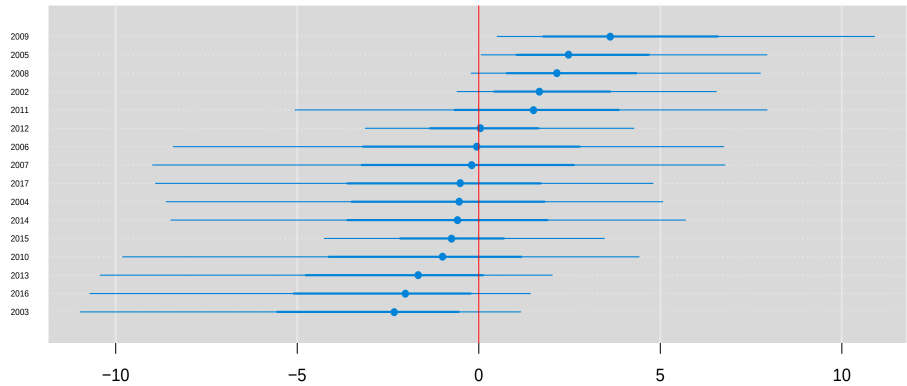
α_i and the year–level random effects γ_t (see Figures 36 and 37) of each model including 95% credible intervals of the estimates, respectively (Fahrmeir et al. 2013, 376–377). While we observe heterogeneity for both country– and year–level random –effects, it is larger for the year–level random–effects (Fahrmeir et al. 2013, 376–407). This is less surprising given that the sample exclusively consists of developed EU member states with a relatively

R–package `mcmcplots` (Curtis 2015).

similar level of economic performance, as shown in Figure 30. Nevertheless, observing this variability for both kinds of random effects illustrates the presence of heterogeneity at not only the country–, but also the year–level in BTSCS data and consequently the importance of modeling it (Fahrmeir et al. 2013, 402). Thus, the regression results confirm the need for considering year–level random intercepts when applying multi-level models on (B)TSCS data.

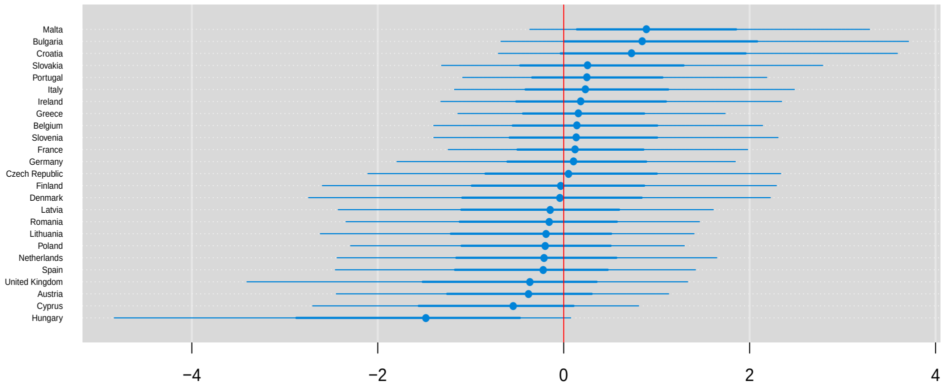


(a) Country–Level Random–Effects α_i

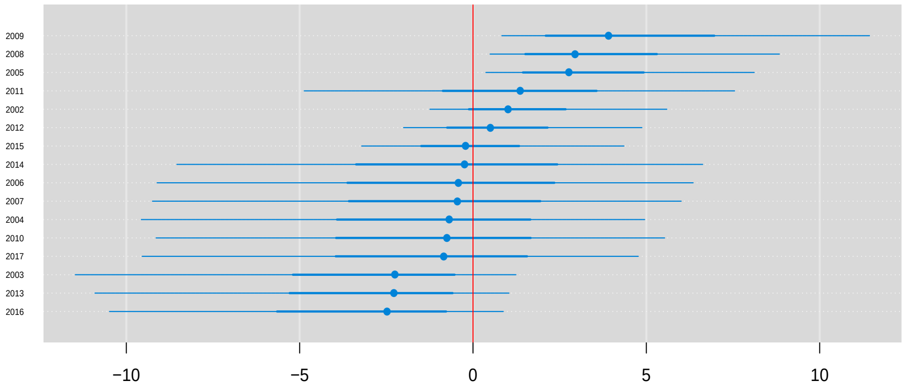


(b) Year–Level Random–Effects γ_t

Figure 36: Random Effects of Model 1 (*Trust in European Commission* $\times$ *Real GDP*).



(a) Country–Level Random–Effects α_i



(b) Year–Level Random–Effects γ_t

Figure 37: Random Effects of Model 2 (*Relative Public Support* × *Real GDP*).

The results of the first and the second model, where the two public support indicators *Trust in European Commission* and *Relative Public Support* have been interacted with *Real GDP*, seemed too uncertain to confirm hypothesis 1 stating that the probability of EDP initiations increases if public support and transparency are high. However, are these results also robust? Figures 43 to 58[7] (see Appendix) show the results of the robustness

[7]Colums 3 and 4 in Table 4 include the coefficient estimates of all robustness check

checks where *Trust in European Commission* and *Relative Public Support* have been interacted with the second transparency measure *GDP Growth*. Furthermore, the third public support indicator, *Positive Image of the EU*, has been interacted with both *Real GDP* and *GDP Growth*, respectively, and six additional models have been run with the alternative specification of the dependent variable.

For the interaction between *Trust in European Commission* and *GDP Growth* (Figures 43 and 44) we can see that the results of the first model still hold: for high levels of support for the Commission and high GDP growth, there is an increase in the probability of an EDP initiation, while the estimation uncertainty of an EDP initiation increases with higher values of trust. However, as for Model 1, the effect of public support does not seem to depend on transparency as higher levels of trust still lead to a higher probability of sanctioning even if transparency is low. For this scenario, the probability of an EDP initiation ranges from 0 to 1 for all values of Trust in European Commission, whereby estimation uncertainty is still very high.

Figures 45 and 46 illustrate the effect of the *Relative Public Support* variable on EDP initiations conditional on transparency measured by GDP growth. From the two plots, it becomes obvious that the results of Model 2 are not robust: when interacted with *GDP Growth*, higher support for the Commission yields a lower probability of EDP initiation when *GDP Growth* is fixed at its maximum, while it increases the probability of an EDP opening if *GDP Growth* is low. The expected probability ranges between 0 and 0.4 for high values of *Relative Public Support* and a maximum of *GDP Growth*, while the uncertainty of an EDP initiation also remains very high for all values of *Relative Public Support* if *GDP Growth* is set to its minimum.

Models 5 and 6 (for coefficient estimates, see Column 5 and 6 in Table 4) consider the fact that it is possible that citizens are skeptical about the Commission itself due to its weak legitimization, but supportive of the EU and its rules in general. Here, public support is measured by the share of respondents expressing having a positive image of the EU and is interacted with *Real GDP* and *GDP Growth*, respectively. Figures 47 and 48 picture the effect of the variable *Positive Image of the EU* on EDP openings conditional on the level of transparency measured by real GDP (high/low). We can see that the higher the share of respondents in a country that express having a positive image of the EU, the more likely it becomes that the Commission will open an EDP against a given country

models. However, since the coefficients are not interpretable, in the robustness checks I will also exclusively focus on the expected values for interpretation and hypothesis testing.

that violated the SGP's rules given that transparency is high. Even if transparency is low, on the other hand, we still see a positive effect of the share of respondents who perceive the EU as positive on the probability of EDP openings, as illustrated by Figure 48 . Once again, the effect of public support seems to be independent of the levels of transparency. However, in both scenarios, the estimation uncertainty of an EDP initiation increases with higher shares of respondents who have a positive impression of the EU.

The same is true for the model where *Positive Image of the EU* has been interacted with *GDP Growth*, as shown in Figures 49 and 50: high values of public support and transparency make sanctioning of SGP violations more likely, while the uncertainty of an EDP initiation increases with a higher share of respondents with a positive opinion about the EU. The positive effect of public support on the probability of EDP openings is still given even without high transparency, but the probability of an EDP initiation ranges between 0 and 1 for all values of the last public support indicator.

For the robustness check with the alternative specification of the dependent variable, six additional models covering all interactions between each public support measure and transparency measure have been run (see Columns 7 to 12, Table 4). Figures 51 to 62 show the results for this additional robustness check. The results of the model including the interaction term *Trust in Commission* × *Real GDP* with *Real GDP* contradicts the predictions of hypothesis 1 as there is no longer a positive effect of high values in *Trust in Commission* on EDP initiations if transparency is high; rather, we see no effect of public support and the uncertainty of an EDP initiation ranges from 0 to 1 for all values of *Trust in Commission*. Somewhat counter-intuitively, high values of *Trust in Commission* increase the probability of an EDP initiation if *Real GDP* equals its minimum, as Figure 52 shows. Nevertheless, the estimation of an EDP initiation remains still very high for high values of public support, as it ranges from 0.1 to 1.

By contrast, the results of the model where *Trust in Commission* is interacted with *GDP Growth* – which are shown in Figures 53 and 54 – once again show that high values of public support and transparency increase the probability of an EDP initiation, albeit with major estimation uncertainty. If *GDP Growth* is low, this positive effect of public support is no longer present, which is different from the model using the first EDP operationalization and highlights the importance of transparency. Nevertheless, the estimation uncertainty of an EDP initiation is also very high for this scenario throughout all values of the public support indicator.

Similar to the first model, the positive effect of high values of *Relative Public Support* on

EDP initiations conditional on high levels of transparency – measured by *Real GDP* – is not robust when using a different specification of EDP initiations, as Figures 55 and 56 show. Here, more support for the Commission in contrast to the national government reduces the probability of an EDP initiation given high transparency, albeit with high estimation uncertainty. The same is true for low values of *Real GDP* and for the interaction of *Relative Public Support* and *GDP Growth* when *GDP Growth* is set to its maximum (see Figure 57). In this case, the probability of an EDP initiation is ranging between 0 and 0.2 if *Relative Public Support* is high. For low values of transparency measured by *GDP Growth* (see Figure 58), a higher support for the Commission than for the national government surprisingly increases the probability of an EDP initiation, but this probability again ranges between 0 and 1. Transparency thus seems to have an effect here, but not in the expected way.

For the models where the last measure of public support, *Positive Image of the EU*, is interacted with each measure of transparency, the results remain robust when a different dependent variable is chosen and in line with hypothesis 1, which is depicted in Figure 59 to 62. Here, the level of transparency again seems to matter less than the level of transparency for the initiation of an EDP, as we observe an increasing probability of sanctioning with rising values of the variable *Positive Image of the EU*, regardless the level of any measure of transparency, although the probability of an EDP initiation also remains very uncertain for this model specification.

Table 5 provides an overview of the results with respect to the effect of public support on EDP initiations on for all model specifications.

Table 5: Overview Results (including Robustness Checks)

Model	Interaction	Dependent Variable	Effect
1	*Trust in Commission* × *Real GDP*	EDP – Opinion)	+
2	*Relative Public Support* × *Real GDP*	EDP – Opinion)	+
3	*Trust in Commission* × *GDP Growth*	EDP – Opinion)	+
4	*Relative Public Support* × *GDP Growth*	EDP – Opinion)	–
5	*Positive Image of the EU* × *Real GDP*	EDP – Opinion)	+
6	*Positive Image of the EU* × *GDP Growth*	EDP – Opinion)	+
7	*Trust in Commission* × *Real GDP*	EDP – Report	–
8.	*Trust in Commission* × *GDP Growth*	EDP – Report	+
9	*Relative Public Support* × *Real GDP*	EDP – Report	–
10	*Relative Public Support* × *GDP Growth*	EDP – Report	–
11	*Positive Image of the EU* × *Real GDP*	EDP – Report	+
12	*Positive Image of the EU* × *GDP Growth*	EDP – Report	+

Finally, Figures 63 to Figure 72 show the caterpillar plots of the country– and year–random effects for all robustness check models. From the caterpillar plots and the estimates of σ_α and σ_γ in the regression tables, we can see that heterogeneity at the year–level is also much larger than heterogeneity at the country level for the robustness check models.

In sum, the results of both the analysis and the robustness checks indicate a positive effect of public support on the probability of EDP initiations as postulated in hypothesis 1 for most models. However, transparency seems to be a less crucial determinant than public support as an increase of this factor made sanctioning SGP violations more likely regardless the level of transparency. Furthermore, the uncertainty of expected probabilities is very high due to few observations in the dataset, which stems from the fact that this investigation focuses exclusively on the EU – which only comprises 28 member states – and the SGP has only been in force since 2002. This data does not provide sufficient information to overrule the uninformative prior. Finally, the results were not robust for the indicators *Trust in European Commission* and *Real GDP* for the different specification of the dependent variable. They were also not robust for *Relative Public Support* – which considered public support for the national government (and implicitly governmental overspending) besides support for the European Commission – when interacted with *GDP Growth* instead of *Real GDP* as a measure of transparency and also for the alternative dependent variable. Hypothesis 1 can therefore be rejected due to high uncertainty of all results and a lack of robustness.

Nevertheless, the results of the analysis support the impression that sanctioning behavior of the Commission is influenced by legitimacy concerns. From a methodological perspective, the results stress the importance of modeling year–level effects in BTSCS data due to high heterogeneity at the year level, which exceeded variation at the country–level. This is particularly true for samples of homogeneous, developed countries like EU or OECD member states.

9 Summary and Discussion

In the preceding chapters, a new theoretical argument based on combining insights from both EU research and Judicial Politics has been developed to explain the ineffectiveness of the European Commission's sanctioning of SGP violations. I have identified the levels of public support and transparency as crucial determinants for the Commission's sanctioning behavior. By doing so, I expand the literature on SGP (in)effectiveness by highlighting the Commission's legitimacy and transparency problem at the member state level which reduces the effectiveness of non–majoritarian fiscal surveillance in the EU. Unfortunately, the empirical test of this argument provides insufficient evidence that high levels of public support and transparency make it more likely that the Commission will sanction SGP violations by initiating an EDP, which is mainly due to the small dataset. This indicates the need for future research to re–test the theoretical argument on a larger dataset to see whether this results change in favor for hypothesis 1.

Nevertheless, most expected values hint at an effect of high levels of public support on EDP initiations. Therefore, this study provides some evidence that the Commission is a strategic actor that only enters conflicts with member state governments if its prospects for success are promising, and that this is not only true for infringement procedures (see, for example, König and Mäder 2014; Fjelstul and Carrubba 2018), but also for EDP procedures.

Given that the dataset used to test the theoretical argument is a small N BTSCS dataset on EU member states for a limited period of time, I propose a Bayesian Multi-level Probit Model with a country– and year–level random intercept (BMLP BTSCS) to model panel heteroskedasticity and contemporaneous correlation. I have provided a test for this model's performance in capturing these BTSCS characteristics to the one of existing techniques in a Monte Carlo experiment. Since the results of the experiment confirmed that the BMLP BTSCS model appropriately accounts for panel heteroskedasticity and contemporaneous correlation, this dissertation also contributes to the methodological literature by providing a broad comparison of BTSCS models and identifying Bayesian multi-level models as an appropriate alternative for analyzing small N BTSCS data.

The findings of this dissertation project hold implications for several areas of political science research and real–world politics.

First, the empirical findings offer a hint that a lack of legitimacy of the Commission hampers the effectiveness of the SGP links the economic performance of the EU back to the discussion of the EU's "democratic deficit" (Follesdal and Hix 2006; Majone 1998; Moravcsik 2002). In particular, the findings of this project question Majone's (1998) claim that "economic integration without political integration is possible only if politics and economics are kept as separate as possible" in two ways (Majone 1998, 5). First, this investigation offered a first insight that it is very difficult to completely separate politics from economics since even NMIs – which are seemingly "unpolitical" – need to be concerned about their legitimacy and hence political power towards elected governments that they are supposed to sanction in case of non–compliance. Second, it might be the ineffectiveness caused by the attempt to separate economics and politics that is likely to slow down economic integration in the future: this is because events like the European sovereign debt crisis, which at least partly stemmed from the ineffectiveness of the SGP, led to increased Euroscepticism and opposition towards further (economic) integration, which has been confirmed by several studies (Braun and Tausendpfund 2014; Clements et al. 2014; Hobolt and Tilley 2016; Kuhn and Stöckel 2014). Moreover, the findings raise the concern of whether the Commission's consideration of public support for their sanctioning decisions could incentivize member state governments to spread Euroscepticism among their voters to continue overspending without any negative consequences.

Second, this study has shown that public support and transparency are not only decisive for the behavior of constitutional courts, but might play a role for other institutions that can, as the court, be classified as NMIs. Besides the European Commission, this includes institutions like independent central banks, regulatory authorities or even International Organizations (Majone 2001, 58). Since the results suggest that public support could influence the behavior of any these NMIs, this raises the question of whether NMIs might suffer from their own kind of "time–inconsistency problem", namely that their behavior is not determined by elections, but varying levels of public support (Majone 2001, 62)[1].

[1]According to Majone (2001), the term "time–inconsistency problem" refers to the situation where "[...] a government's optimal long–run policy differs from its optimal short–run policy [...]", which often incentivizes the government to deviate from its long–term commit-

Third, methodologically, the higher amount of heterogeneity among years than among countries in the data used here stressed the importance of modeling not only country–, but also year–level random effects when using Multi-level models on (B)TSCS data. This is particularly true for research on the EU or the OECD, where the samples comprise very similar countries that are closely linked to each other politically and economically. This structure makes it likely that there is both contemporaneous correlation and more variance between years than across countries present.

Fourth, in terms of implications for EU politics, governments and EU officials need to particularly address the problem of lacking legitimacy of the Commission to improve the functioning of the SGP and consequently to prevent high debts and deficits in the future. A possible solution would be the establishment of competitive European Parliamentary Elections, with the result of a supranational European government that is accountable to the parliament and hence to the voters (Hix and Høyland 2011, 132–137). Such an institution would possess the necessary amount of legitimacy and political power to sanction any member state government that violates the SGP's reference values, thereby ensuring the proper function of the Eurozone in particular and the EU in general. Admittedly, this is unfortunately unlikely to happen in the near future due to the above mentioned rising Euroscepticism caused by the European sovereign debt crisis.

Finally, there naturally several possibilities exist to improve and extend the analysis. First and foremost, the dataset used for this analysis is rather small as it comprises less than 30 countries and less than 20 years, which contributed of to excessive estimation uncertainty of the results (King et al. 2000, 355). Re–running the analysis with a sample that covers a longer period of time, ideally with more variance on the dependent variable, would make the current results more reliable. It would also be possible to split–up the sample into countries that violated the deficit rule and countries that violated the debt rule to ascertain whether the Commission's behavior differs with the SGP rule that has been violated, particularly since the deficit criterion has been considered as more important in the past (Bauer and Becker 2014, 220–227). Methodologically, it would be possible to account for serial correlation in the individual error terms to further avoid overconfidence

ments Majone 2001, 62). The time–inconsistency problem is most prominent in monetary policy, where governments are assumed to have an inflationary bias to stimulate the economy before an election, which can lead to unexpected inflation (Eijffinger and De Haan 1996, 5–7). As a solution to this problem, monetary policy is delegated to an independent central bank (Eijffinger and De Haan 1996, 5–7).

in the regression results, as suggested by Pang (2010) (Pang 2010, 470–475). In order to test this model's performance, a more extensive Monte Carlo experiment, probably even with varying levels of panel heteroskedasticity, contemporaneous correlation and serial correlation, could be conducted (Shor et al. 2007, 173). Furthermore, if the researcher suspects a country– (or year) –varying effect of some independent variables, the model could also be extended by adding a random–slope coefficient (Beck and Katz 2007, 183–184; Gelman and Hill 2006, 237–238; Shor et al. 2007, 176).

I have chosen to use non–informative priors for this analysis. However, Bayesian models generally allow for including additional information available in the analysis by using informative priors (Gelman and Hill 2006, 392–393). If researchers have clearer expectations about the effect of some variables on EDP initiations in the future, including these information in the regression model via informative priors for the β–coefficients may lead to further improvements of the results. Another possibility would be the application of a mixed–methods design that would allow the researcher to gain a more in–depth understanding of the Commission's sanctioning behavior by combining quantitative with qualitative methods such as case studies (Bäck and Dumont 2007, 469; Teddlie and Tashakkori 2009).

10 Concluding Remarks

This study has focused on the political reasons for the Commission's hesitant sanctioning behavior of SGP violations in particular and the ineffectiveness of supranational non–majoritarian economic governance in general. Theoretically, it combines arguments from EU research and judicial politics on the strategic nature of NMI decision–making and expands the literature on SGP research by the argument that the lack of legitimacy and transparency explain the European Commission's ineffective sanctioning of SGP violations. Empirically, it adds to this literature by providing the first quantitative analysis that tests the impact of public support and transparency on EDP initiations by the Commission.

In terms of methodology, the dataset used for the quantitative analysis conducted here comprised data for 25 EU states for 2002 to 2017, which classifies them as BTSCS data. Analyzing such data is methodologically challenging, as they comprise information on a few countries for a limited time period that are both politically and economically highly intertwined with each other. This requires to simultaneously account for small sample size, contemporaneous correlation and panel heteroskedasticity and raises the question on which method is most suitable to address all these characteristics. For this purpose, a BMLP BTSCS model with a country– and year–specific random intercept has been introduced and tested compared with the performance of existing methods for BTSCS data analysis in a Monte Carlo experiment. The results of the experiments showed that the BMLP BTSCS model yields unbiased estimates and correct standard errors for small sample sizes. This dissertation therefore also contributes to the methodological literature on BTSCS data analysis by identifying a suitable alternative to existing BTSCS modeling techniques.

The results of the investigation about the determinants of EDP openings by the European Commission once again suggest that there is the need for further investigation of whether the Commission's (sanctioning) behavior might be driven by strategic considerations (König and Mäder 2014; Köhler et al. 2018). It has been demonstrated that the Commission is more likely to initiate an EDP if public support *and* transparency about a country's economic situation are sufficiently high, but the results are not robust if support for the national government is considered and estimation uncertainty is high due to the small dataset. Nevertheless, an important implication of this finding is that the Commission's legitimacy and transparency problem inherent in the its institutional design might be able to eliminate the SGP's sanctioning mechanism as they could induce the Commission to only act if it considers both problems to be solved.

There are a number in which this investigation could be improved. As previously mentioned, this analysis' dataset is rather small due to limited data availability. Furthermore, there is very little variation on the dependent variable across countries as most EU member states have been at least once experienced an EDP and since 2013, the Commission has not opened a EDP against any country. These drawbacks of the current data make it necessary to re–run the analysis in the future with a dataset that comprises both more observations and variance. Another aspect is that violating the debt criterion has in the past been considered as less crucial for sanctioning than a violation of the deficit reference value. In order to ascertain whether the Commission's sanctioning behavior differs for the violation of either SGP rule, one could run the analysis on two different samples of countries where the first group violated the debt rule and the second one the deficit rule only.

Besides accounting for panel heteroskedasticity and contemporaneous correlation, the BMLP model introduced in this study could be extended by considering serial correlation in the individual–level errors. The modification of the current method would then require a more extensive Monte Carlo experiment where the degree of panel heteroskedasticity, contemporaneous correlation and serial correlation is allowed to vary. The inclusion of a random–slope would further allow the researcher to detect country– or year–varying effects of some independent variables. Moreover, adding information about the effect of some variables on EDP initiations that might be available in the future via the use of informative priors may also yield more precise results.

The focus of this dissertation has clearly been on the use of quantitative methods. Applying a mixed–methods design where quantitative methods are combined with qualitative methods like case studies would allow for deeper insights of the underlying causal mechanism that explains the Commission's sanctioning decisions.

As a concluding remark, the findings of this study suggest that the "democratic deficit" of the Commission that stems from its lack of legitimacy and that might contribute to the ineffectiveness of the SGP requires further investigation. In particular, previous claims on the separability of political and economic integration need to be reevaluated. For the future of the EMU, it is crucial that particularly voters – rather than governments or EU elites – understand this relationship, given that their interests are the driving force of successful EU integration.

11 Appendix

11.1 Monte Carlo Experiment

11.1.1 Variance-Covariance Matrix of Errors for BTSCS Data

According to Wallace and Hussain (1969), the variance-covariance matrix $\sum$ is defined as

$$\sum = \sigma_\omega^2 I_{NT} + \sigma_{\alpha_i}^2 D + \sigma_\gamma^2 B \tag{11.1}$$

Beck and Katz 1995, 636; Carsey and Harden 2013, 107-108; Shor et al. 2007, 172-173; Singer and Willett 2003, 249-250; Wallace and Hussain 1969, 56-58.

I_{NT} is an $NT \times NT$ identity matrix denoted as

$$I_{NT} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

D being a $NT \times NT$ block diagonal matrix given by

$$D = \begin{bmatrix} J_T & 0 & 0 & \dots & 0 \\ 0 & J_T & 0 & \dots & 0 \\ 0 & 0 & J_T & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & J_T \end{bmatrix};$$

where J_T is a $T \times T$ matrix of ones;

$$J_T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix};$$

and B being a $NT \times NT$ block matrix equal to

$$B = \begin{bmatrix} I_T & I_T & I_T & \dots & I_T \\ I_T & I_T & I_T & \dots & I_T \\ I_T & I_T & I_T & \dots & I_T \\ \vdots & \vdots & \vdots & \ddots & I_T \\ I_T & I_T & I_T & I_T & I_T \end{bmatrix};$$

where I_T is a $T \times T$ identity matrix

$$I_T = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Wallace and Hussain 1969, 56-58. For D and B, there are N rows and columns of the block matrices J_T and I_T, respectively Beck and Katz 1995, 636; Carsey and Harden 2013, 107-108; Shor et al. 2007, 172-173; Singer and Willett 2003, 249-250; Wallace and Hussain 1969, 56-58.

For $N = 3$ (= three countries) and $T = 3$ (= three years), the covariance matrix $\sum$ can be calculated as follows:

$$\sum = \sigma_\omega^2 \begin{bmatrix} 1&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1 \end{bmatrix} + \sigma_{\alpha_i}^2 \begin{bmatrix} 1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&1&1&1 \end{bmatrix} + \sigma_\gamma^2 \begin{bmatrix} 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1 \end{bmatrix}$$

$$\sum = \begin{bmatrix} \sigma_\omega^2&0&0&0&0&0&0&0&0\\ 0&\sigma_\omega^2&0&0&0&0&0&0&0\\ 0&0&\sigma_\omega^2&0&0&0&0&0&0\\ 0&0&0&\sigma_\omega^2&0&0&0&0&0\\ 0&0&0&0&\sigma_\omega^2&0&0&0&0\\ 0&0&0&0&0&\sigma_\omega^2&0&0&0\\ 0&0&0&0&0&0&\sigma_\omega^2&0&0\\ 0&0&0&0&0&0&0&\sigma_\omega^2&0\\ 0&0&0&0&0&0&0&0&\sigma_\omega^2 \end{bmatrix} + \begin{bmatrix} \sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0&0&0&0\\ \sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0&0&0&0\\ \sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0&0&0&0\\ 0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0\\ 0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0\\ 0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&0&0&0\\ 0&0&0&0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2\\ 0&0&0&0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2\\ 0&0&0&0&0&0&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2&\sigma_{\alpha_i}^2 \end{bmatrix} + \begin{bmatrix} \sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0\\ 0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0\\ 0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2\\ \sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0\\ 0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0\\ 0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2\\ \sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0\\ 0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0\\ 0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2&0&0&\sigma_\gamma^2 \end{bmatrix}$$

$$\sum = \begin{bmatrix} \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2 & \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & 0 & 0\\ \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2 & 0 & \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & 0\\ \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2\\ \sigma_\gamma^2 & 0 & 0 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2 & \sigma_\gamma^2 & 0 & 0\\ 0 & \sigma_\gamma^2 & 0 & \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2 & 0 & \sigma_\gamma^2 & 0\\ 0 & 0 & \sigma_\gamma^2 & \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2\\ \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & 0 & 0 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2\\ 0 & \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & 0 & \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 & \sigma_{\alpha_i}^2\\ 0 & 0 & \sigma_\gamma^2 & 0 & 0 & \sigma_\gamma^2 & \sigma_{\alpha_i}^2 & \sigma_{\alpha_i}^2 & \sigma_\omega^2+\sigma_{\alpha_i}^2+\sigma_\gamma^2 \end{bmatrix}$$

11.1.2 Monte Carlo Experiment Results: Tables

Table 6: Monte Carlo Experiment: Absolute Bias

N	T	Probit	B & K	Freq. MLP	Freq. MLP BTSCS	BMLP	BMLP BTSCS
10	10	0.27	0.26	0.30	0.30	0.28	0.29
10	20	0.17	0.14	0.27	0.26	0.12	0.10
10	30	0.17	0.15	0.26	0.25	0.06	0.03
10	40	0.18	0.16	0.27	0.26	0.10	0.06
10	50	0.17	0.16	0.26	0.26	0.09	0.08
20	10	0.20	0.19	0.27	0.27	0.11	0.10
20	20	0.20	0.18	0.28	0.27	0.17	0.14
20	30	0.17	0.14	0.27	0.26	0.12	0.09
20	40	0.17	0.17	0.27	0.26	0.11	0.09
20	50	0.16	0.15	0.27	0.26	0.10	0.08
30	10	0.17	0.16	0.26	0.26	0.07	0.04
30	20	0.17	0.16	0.26	0.26	0.10	0.08
30	30	0.19	0.18	0.27	0.27	0.12	0.10
30	40	0.15	0.14	0.26	0.26	0.10	0.07
30	50	0.16	0.14	0.26	0.26	0.09	0.06
40	10	0.19	0.17	0.27	0.26	0.12	0.09
40	20	0.15	0.14	0.27	0.26	0.10	0.07
40	30	0.15	0.14	0.27	0.26	0.10	0.08
40	40	0.16	0.14	0.26	0.26	0.09	0.06
40	50	0.19	0.18	0.27	0.26	0.12	0.10
50	10	0.14	0.13	0.26	0.26	0.09	0.05
50	20	0.16	0.15	0.27	0.26	0.12	0.10
50	30	0.17	0.17	0.27	0.26	0.11	0.09
50	40	0.18	0.18	0.27	0.27	0.13	0.11
50	50	0.17	0.16	0.27	0.26	0.11	0.08

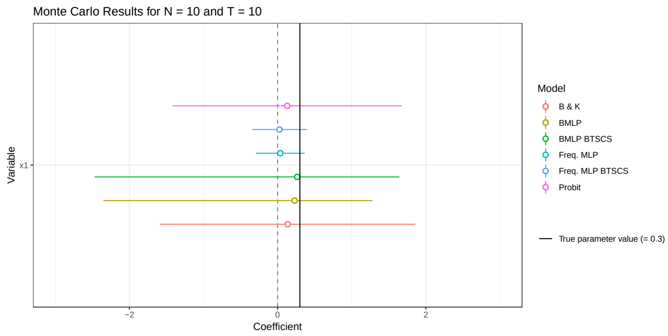
Table 7: Monte Carlo Experiment: Mean Squared Error

N	T	Probit	B & K	Freq. MLP	Freq. MLP BTSCS	BMLP	BMLP BTSCS
10	10	0.79	1.05	0.12	0.16	1.41	2.86
10	20	0.38	0.51	0.09	0.09	0.59	0.86
10	30	0.21	0.28	0.08	0.07	0.29	0.35
10	40	0.19	0.24	0.08	0.08	0.25	0.30
10	50	0.16	0.20	0.08	0.08	0.20	0.24
20	10	0.37	0.43	0.09	0.09	0.53	0.71
20	20	0.19	0.21	0.08	0.08	0.25	0.30
20	30	0.14	0.16	0.08	0.07	0.17	0.19
20	40	0.10	0.11	0.07	0.07	0.11	0.13
20	50	0.09	0.10	0.07	0.07	0.09	0.11
30	10	0.23	0.26	0.08	0.08	0.37	0.46
30	20	0.13	0.14	0.08	0.07	0.18	0.19
30	30	0.11	0.11	0.08	0.07	0.12	0.13
30	40	0.07	0.07	0.07	0.07	0.08	0.08
30	50	0.06	0.06	0.07	0.07	0.07	0.07
40	10	0.17	0.19	0.08	0.08	0.27	0.32
40	20	0.10	0.11	0.07	0.07	0.13	0.15
40	30	0.07	0.08	0.07	0.07	0.09	0.09
40	40	0.06	0.06	0.07	0.07	0.06	0.07
40	50	0.06	0.06	0.07	0.07	0.06	0.06
50	10	0.14	0.15	0.08	0.07	0.21	0.25
50	20	0.08	0.09	0.07	0.07	0.09	0.10
50	30	0.07	0.07	0.07	0.07	0.07	0.08
50	40	0.06	0.07	0.07	0.07	0.06	0.07
50	50	0.05	0.05	0.07	0.07	0.04	0.04

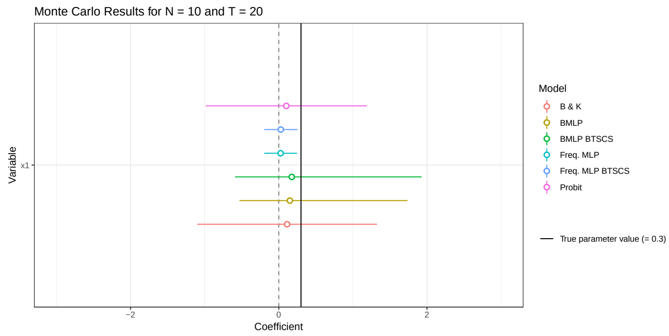
Table 8: Monte Carlo Experiment: Standard Errors

N	T	Probit	B & K	Freq. MLP	Freq. MLP BTSCS	BMLP	BMLP BTSCS
10	10	0.06	0.12	0.02	0.07	0.12	0.39
10	20	0.03	0.08	0.01	0.03	0.07	0.14
10	30	-0.03	0.00	0.00	0.01	-0.02	-0.02
10	40	0.01	0.04	0.01	0.03	0.02	0.03
10	50	0.02	0.03	0.01	0.03	0.02	0.03
20	10	0.03	0.05	0.01	0.01	0.03	0.06
20	20	0.00	0.01	0.00	0.01	0.00	0.01
20	30	0.03	0.04	0.00	0.01	0.01	0.01
20	40	-0.01	-0.00	-0.00	0.00	-0.02	-0.02
20	50	0.02	0.03	0.00	0.01	-0.00	-0.00
30	10	0.01	0.02	0.01	0.01	0.04	0.06
30	20	0.01	0.01	0.01	0.01	0.02	0.01
30	30	0.02	0.01	0.00	0.01	0.02	0.01
30	40	-0.01	-0.00	-0.00	0.00	-0.01	-0.01
30	50	-0.00	-0.00	0.00	0.01	0.00	0.00
40	10	-0.01	-0.00	0.00	0.01	0.02	0.03
40	20	0.01	0.02	0.01	0.01	0.01	0.02
40	30	0.01	0.01	0.00	0.01	0.01	0.01
40	40	0.01	0.01	0.00	0.00	0.01	0.01
40	50	-0.01	-0.00	-0.00	0.00	-0.00	-0.01
50	10	0.00	0.01	0.00	0.01	0.02	0.03
50	20	-0.01	0.00	-0.00	0.00	-0.02	-0.01
50	30	0.00	0.01	0.00	0.01	0.01	0.01
50	40	0.00	0.01	0.00	0.01	0.01	0.02
50	50	-0.00	-0.00	-0.00	-0.00	-0.02	-0.02

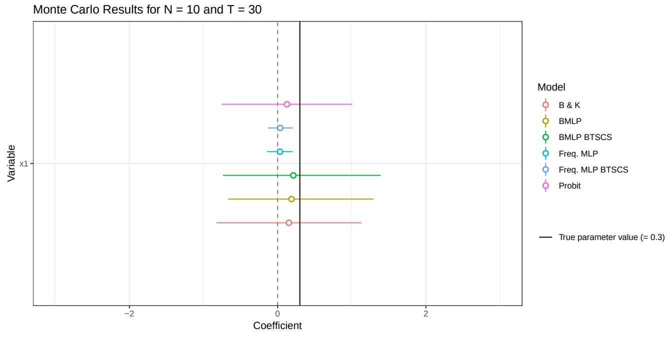
11.1.3 Monte Carlo Experiment Results: Graphs



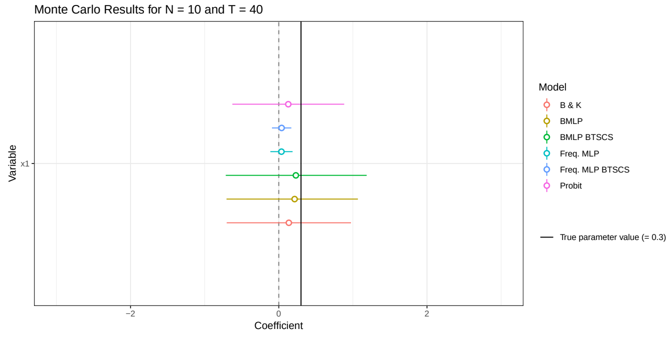
(a) $N = 10$ and $T = 10$



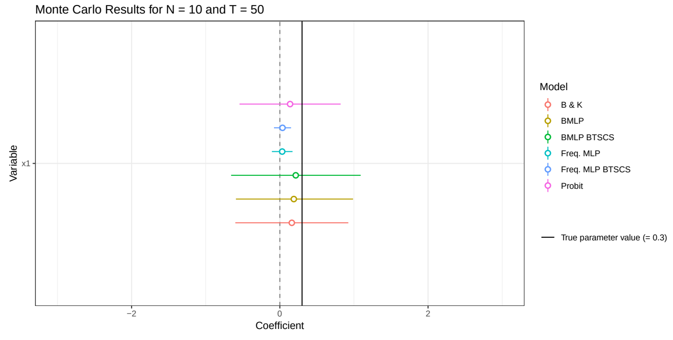
(b) $N = 10$ and $T = 20$



(c) $N = 10$ and $T = 30$

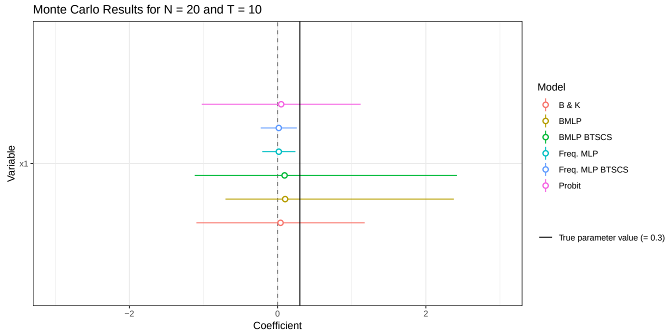


(d) $N = 10$ and $T = 40$

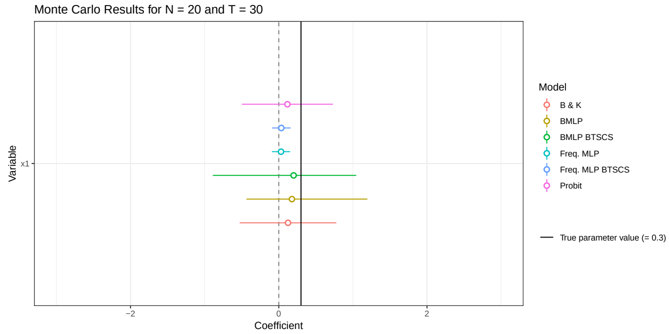


(e) $N = 10$ and $T = 50$

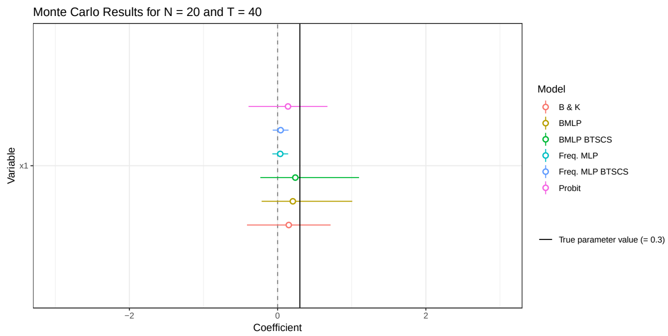
Figure 38: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 10$ and T Ranging from 10 to 50.



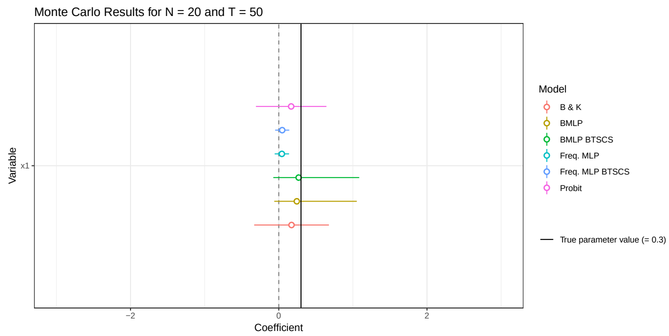
(a) $N = 20$ and $T = 10$



(b) $N = 20$ and $T = 30$

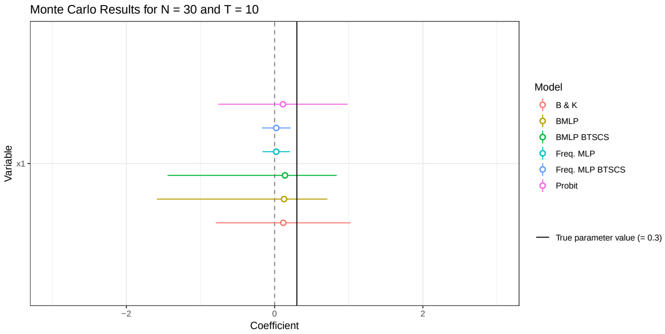


(c) $N = 20$ and $T = 40$

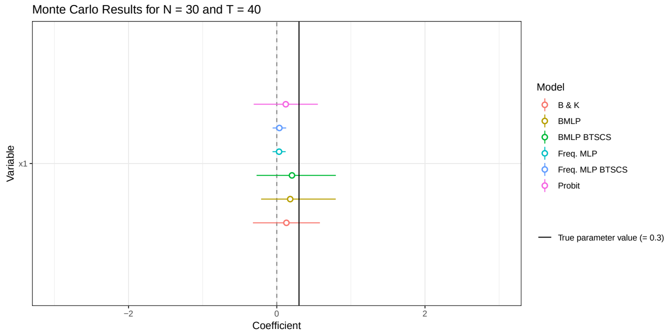


(d) $N = 20$ and $T = 50$

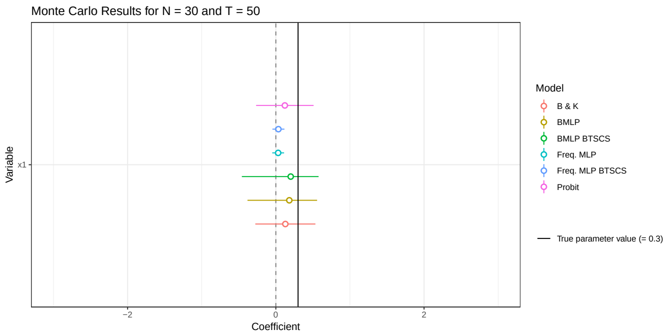
Figure 39: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 20$ and T Ranging from 10 to 50 (Except for $N = 20$ and $T = 20$).



(a) $N = 30$ and $T = 10$

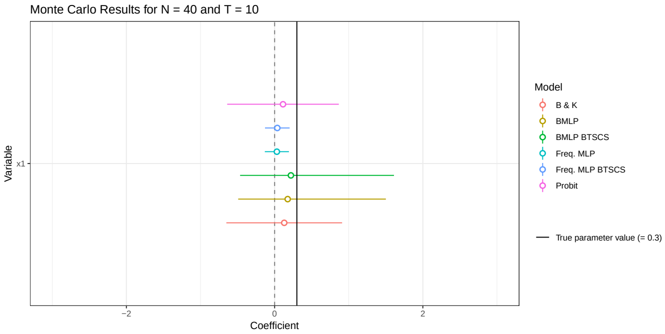


(b) $N = 30$ and $T = 40$

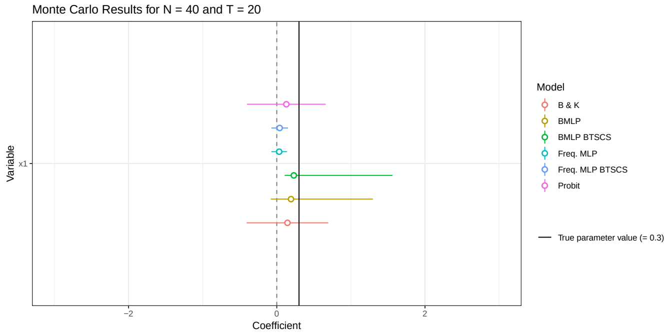


(c) $N = 30$ and $T = 50$

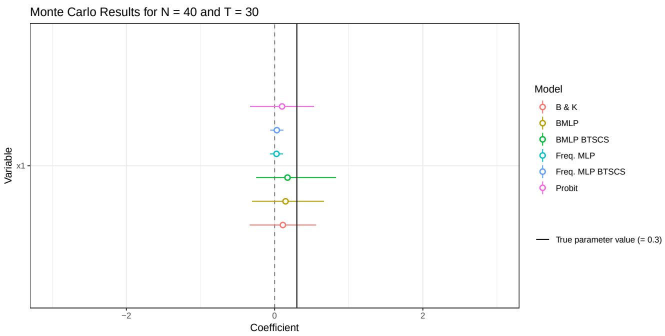
Figure 40: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 30$ and T Ranging from 10 to 50 (Except for $N = 30$ and $T = 20/30$).



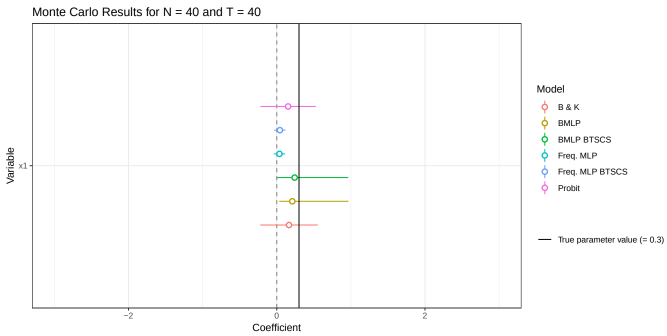
(a) $N = 40$ and $T = 10$



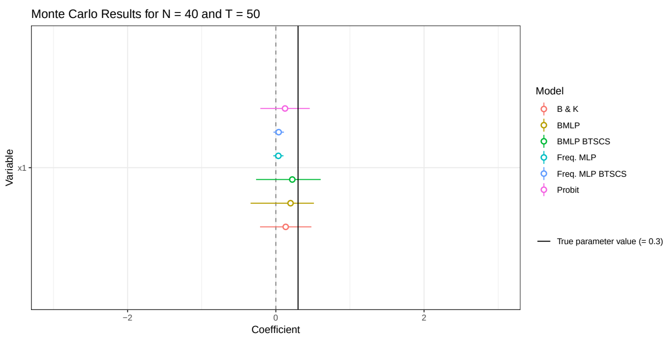
(b) $N = 40$ and $T = 20$



(c) $N = 40$ and $T = 30$

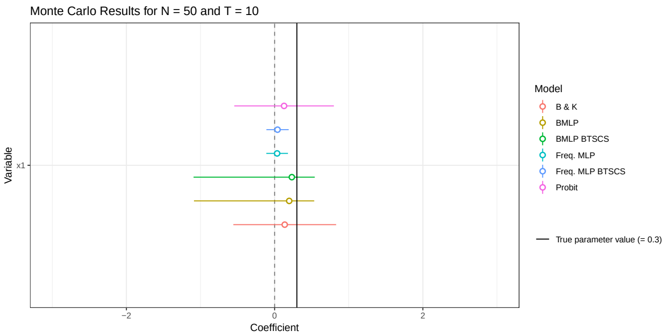


(d) $N = 40$ and $T = 40$

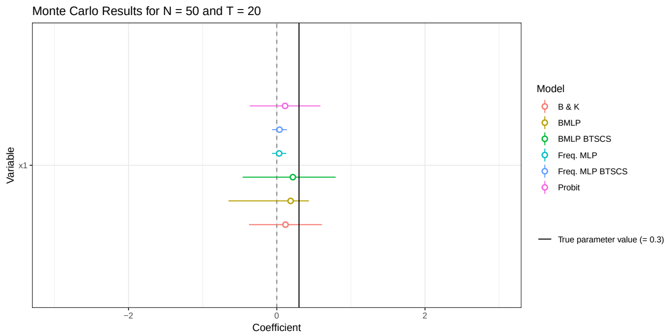


(e) $N = 40$ and $T = 50$

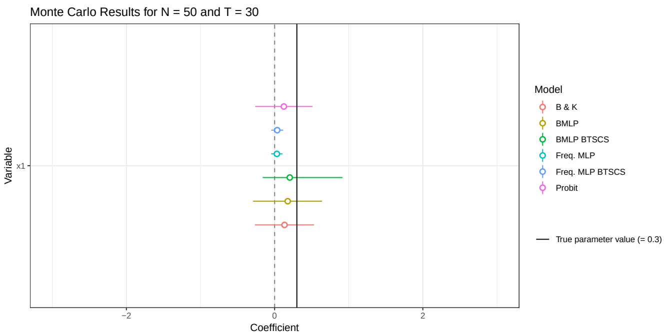
Figure 41: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 40$ and T Ranging from 10 to 50.



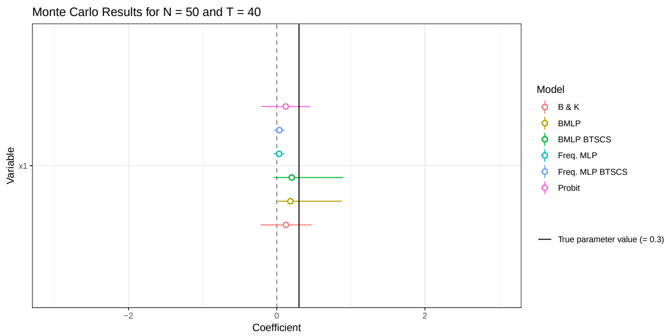
(a) $N = 50$ and $T = 10$



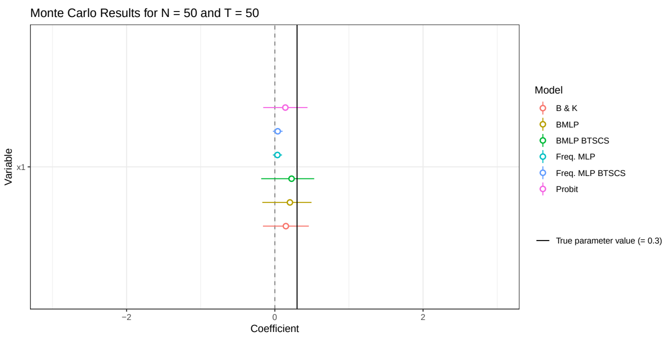
(b) $N = 50$ and $T = 20$



(c) $N = 50$ and $T = 30$



(d) $N = 50$ and $T = 40$



(e) $N = 50$ and $T = 50$

Figure 42: Coefficient Estimates and 95% Confidence/Credible Intervals for $N = 50$ and T Ranging from 10 to 50.

11.2 Descriptive Statistics - Tables

Table 9: Number of SGP Violations per Country

	Country	Number of SGP Violations
1	Austria	7
2	Belgium	7
3	Bulgaria	2
4	Croatia	2
5	Cyprus	6
6	Czech Republic	2
7	Denmark	1
8	Finland	2
9	France	14
10	Germany	8
11	Greece	12
12	Hungary	8
13	Ireland	5
14	Italy	12
15	Latvia	2
16	Lithuania	3
17	Malta	6
18	Netherlands	5
19	Poland	3
20	Portugal	12
21	Romania	2
22	Slovakia	2
23	Slovenia	5
24	Spain	7
25	United Kingdom	9

Table 10: Number of SGP Violations per Year

	Year	Number of SGP Violations
1	2002	3
2	2003	6
3	2004	8
4	2005	7
5	2006	4
6	2007	2
7	2008	16
8	2009	21
9	2010	12
10	2011	14
11	2012	13
12	2013	13
13	2014	14
14	2015	6
15	2016	4
16	2017	1

Table 11: Number of EDPs (Opinion) per Country

	Country	Number of EDPs
1	Austria	1
2	Belgium	1
3	Bulgaria	1
4	Croatia	1
5	Cyprus	1
6	Czech Republic	1
7	Denmark	0
8	Finland	0
9	France	2
10	Germany	2
11	Greece	2
12	Hungary	0
13	Ireland	1
14	Italy	2
15	Latvia	1
16	Lithuania	1
17	Malta	3
18	Netherlands	1
19	Poland	1
20	Portugal	2
21	Romania	1
22	Slovakia	1
23	Slovenia	1
24	Spain	1
25	United Kingdom	1

Table 12: Number of EDPs per Country (Report)

	Country	Number of EDPs
1	Austria	1
2	Belgium	0
3	Bulgaria	1
4	Croatia	1
5	Cyprus	1
6	Czech Republic	1
7	Denmark	0
8	Finland	2
9	France	2
10	Germany	2
11	Greece	1
12	Hungary	0
13	Ireland	1
14	Italy	3
15	Latvia	1
16	Lithuania	1
17	Malta	3
18	Netherlands	0
19	Poland	1
20	Portugal	2
21	Romania	1
22	Slovakia	1
23	Slovenia	1
24	Spain	1
25	United Kingdom	1

Table 13: Number of EDPs (Opinion) per Year

	Year	Number of EDPs
1	2002	0
2	2003	2
3	2004	3
4	2005	2
5	2006	0
6	2007	0
7	2008	1
8	2009	18
9	2010	1
10	2011	0
11	2012	0
12	2013	2
13	2014	0
14	2015	0
15	2016	0
16	2017	0

Table 14: Number of EDPs (Report) per Year

	Year	Number of EDPs
1	2002	1
2	2003	1
3	2004	3
4	2005	2
5	2006	0
6	2007	0
7	2008	1
8	2009	15
9	2010	1
10	2011	0
11	2012	0
12	2013	2
13	2014	1
14	2015	1
15	2016	1
16	2017	0

Table 15: Descriptive Statistics: Dependent and Independent Variables

	EDP - Opinion	EDP - Report	Net Lending/Borrowing	Debt
1	Min. :0.0000	Min. :0.0000	Min. :-32.000	Min. : 12.40
2	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.: -6.325	1st Qu.: 62.98
3	Median :0.0000	Median :0.0000	Median : -4.200	Median : 79.75
4	Mean :0.2014	Mean :0.2014	Mean : -5.325	Mean : 80.38
5	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.: -3.400	3rd Qu.:100.42
6	Max. :1.0000	Max. :1.0000	Max. : 0.700	Max. :180.80

Table 16: Descriptive Statistics: Independent Variables (1)

	Trust in European Commission	Trust in nat. Government	Positive Image of the EU
1	Min. :0.1703	Min. :0.08107	Min. :0.1620
2	1st Qu.:0.3699	1st Qu.:0.19003	1st Qu.:0.3168
3	Median :0.4562	Median :0.28799	Median :0.4041
4	Mean :0.4510	Mean :0.29502	Mean :0.4031
5	3rd Qu.:0.5489	3rd Qu.:0.37717	3rd Qu.:0.5045
6	Max. :0.6617	Max. :0.65000	Max. :0.6492

Table 17: Descriptive Statistics: Independent Variables (2)

	Relative Public Support	Real GDP	GDP Growth
1	Min. :-0.11223	Min. : 6.063	Min. :-14.80000
2	1st Qu.: 0.07333	1st Qu.: 156.957	1st Qu.: -1.10000
3	Median : 0.15894	Median : 328.575	Median : 0.75000
4	Mean : 0.15596	Mean : 971.039	Mean : 0.07222
5	3rd Qu.: 0.25681	3rd Qu.:1966.574	3rd Qu.: 1.80000
6	Max. : 0.40210	Max. :3752.366	Max. : 8.50000

Table 18: Descriptive Statistics: Independent Variables (3)

	Trust in European Commission × Real GDP	Relative Public Support × Real GDP	Positive Image of the EU × Real GDP
1	Min. : 3.577	Min. :-260.052	Min. : 2.739
2	1st Qu.: 74.276	1st Qu.: 8.218	1st Qu.: 50.211
3	Median : 170.478	Median : 44.508	Median : 137.646
4	Mean : 384.397	Mean : 110.891	Mean : 375.534
5	3rd Qu.: 662.600	3rd Qu.: 149.579	3rd Qu.: 662.494
6	Max. :1575.826	Max. : 665.686	Max. :1737.222

Table 19: Descriptive Statistics: Independent Variables (4)

	Trust in European Commission × GDP Growth	Relative Public Support × GDP Growth	Positive Image of the EU × GDP Growth
1	Min. :-7.25876	Min. :-4.77419	Min. :-6.67120
2	1st Qu.:-0.38979	1st Qu.:-0.16465	1st Qu.:-0.34128
3	Median : 0.29942	Median : 0.02324	Median : 0.28337
4	Mean : 0.09728	Mean :-0.02494	Mean : 0.08717
5	3rd Qu.: 0.78261	3rd Qu.: 0.25527	3rd Qu.: 0.75820
6	Max. : 5.19798	Max. : 2.62112	Max. : 5.38801

Table 20: Descriptive Statistics: Independent Variables (5)

	Powerful Member State	Non-Compliance	Eurozone Member
1	Min. :1.000	Min. : 1.00	Min. :1.000
2	1st Qu.:1.000	1st Qu.:11.00	1st Qu.:1.000
3	Median :1.000	Median :14.00	Median :2.000
4	Mean :1.319	Mean :13.33	Mean :1.743
5	3rd Qu.:2.000	3rd Qu.:16.00	3rd Qu.:2.000
6	Max. :2.000	Max. :22.00	Max. :2.000

Table 21: Descriptive Statistics: Independent Variables (6)

	Ex. Election	Leg. Election	Election
1	Min. :1.000	Min. :1.000	Min. :1.000
2	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000
3	Median :1.000	Median :1.000	Median :1.000
4	Mean :1.076	Mean :1.257	Mean :1.299
5	3rd Qu.:1.000	3rd Qu.:2.000	3rd Qu.:2.000
6	Max. :2.000	Max. :2.000	Max. :2.000

11.3 Robustness Checks

11.3.1 Expected Values Simulation

Alternative Specification of Independent Variables

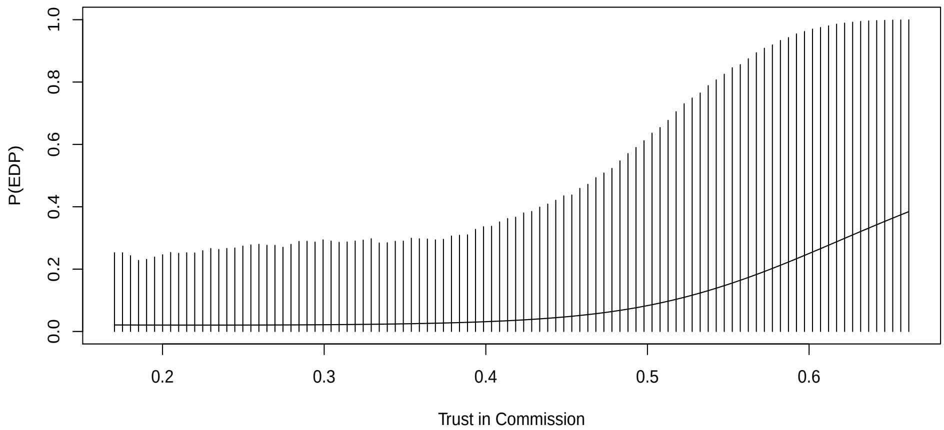


Figure 43: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = high)

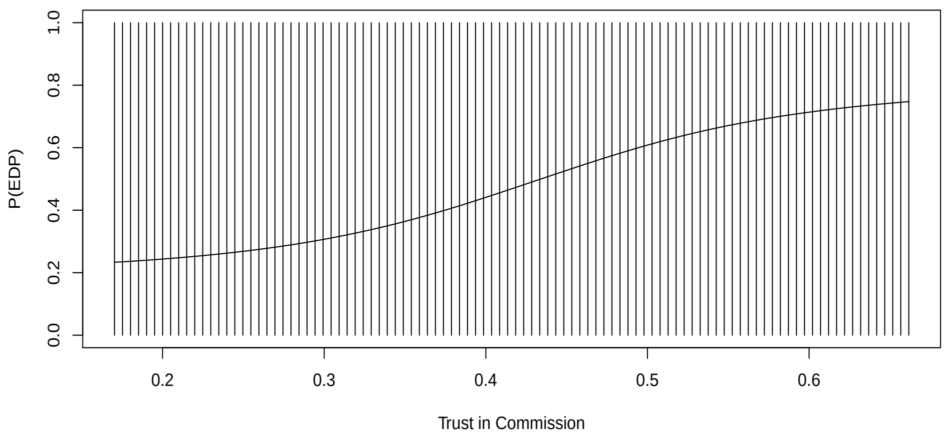


Figure 44: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = low)

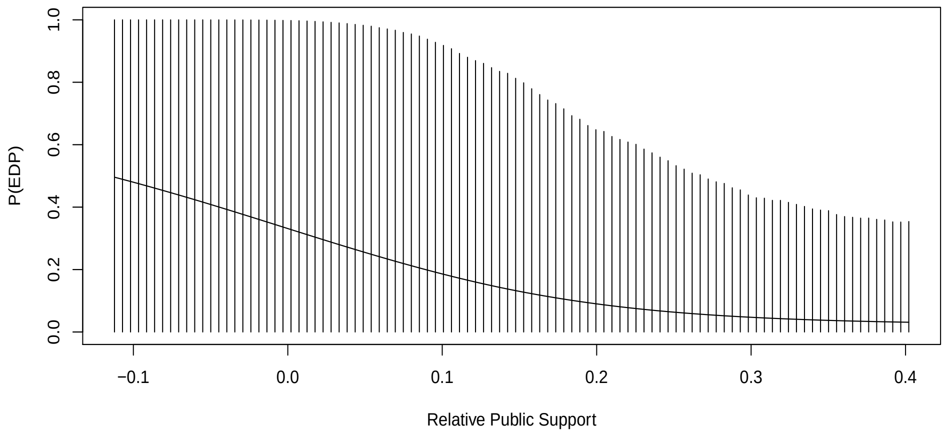


Figure 45: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = high)

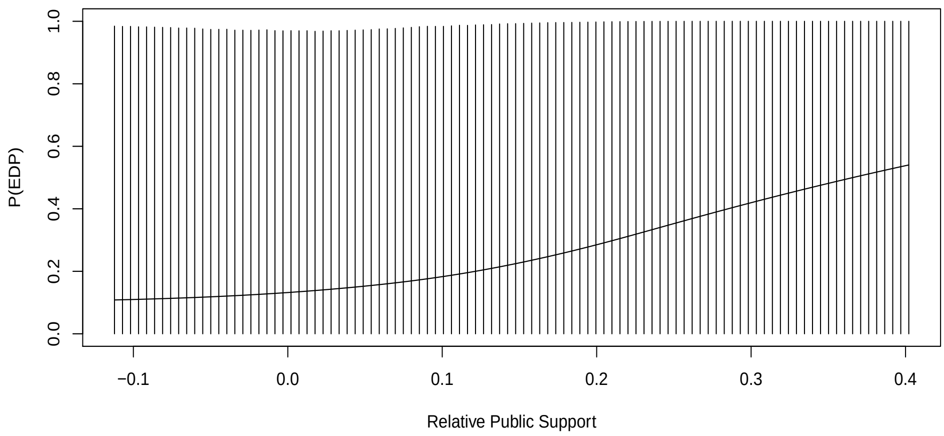


Figure 46: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = low)

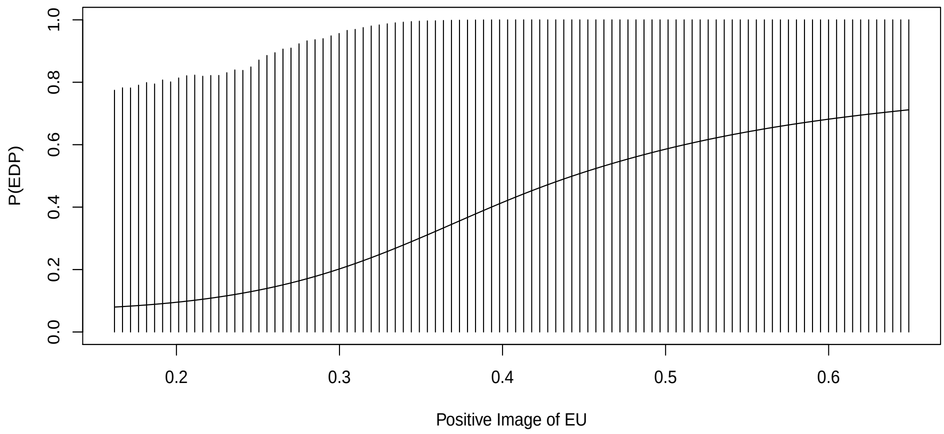


Figure 47: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *Real GDP* on EDP (Opinion) Initiations (*Real GDP* = high)

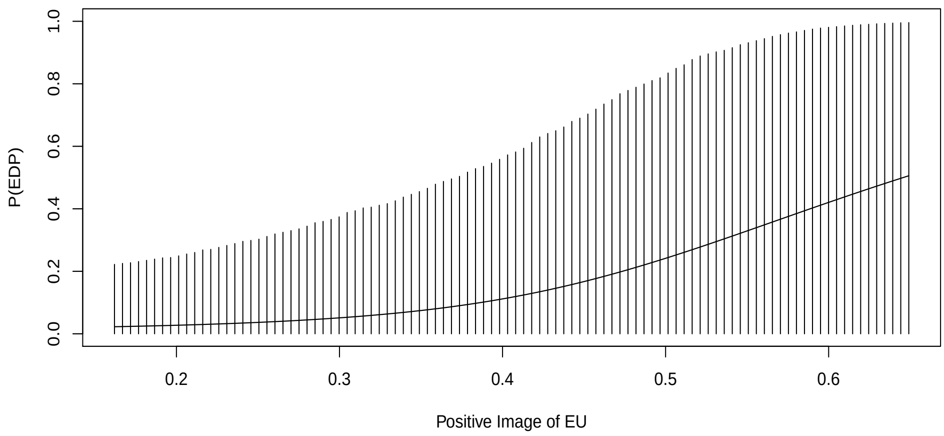


Figure 48: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *Real GDP* on EDP (Opinion) Initiations (*Real GDP* = low)

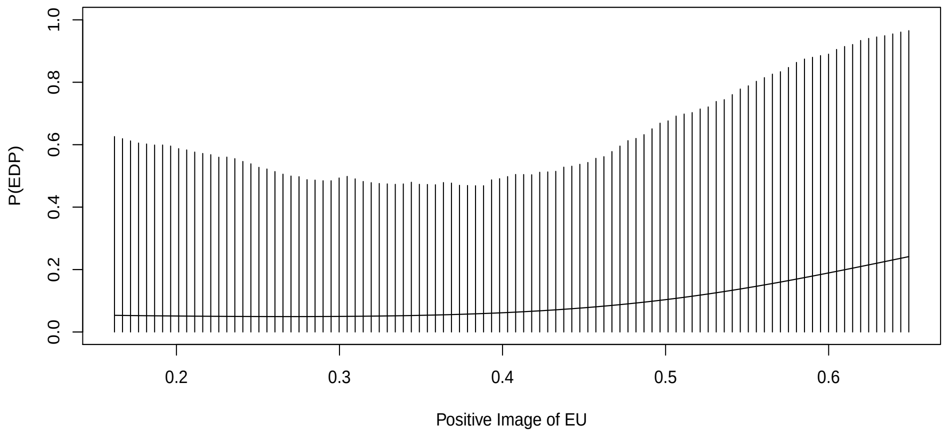


Figure 49: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = high)

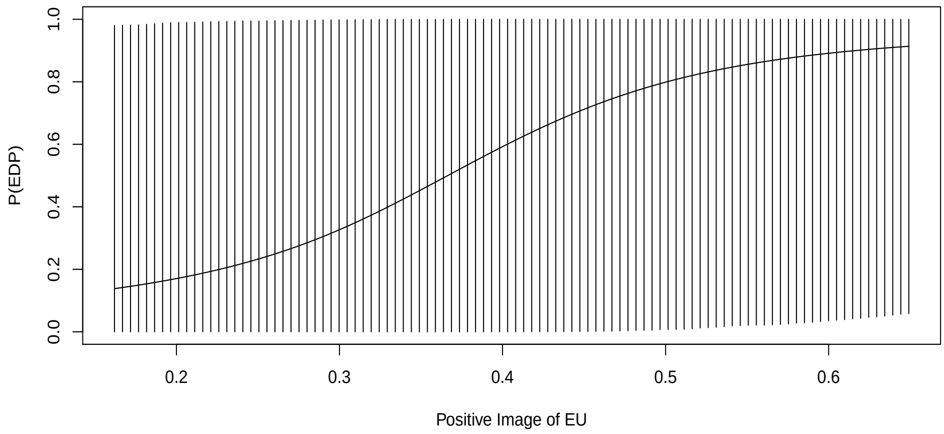


Figure 50: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *GDP Growth* on EDP (Opinion) Initiations (*GDP Growth* = low)

Alternative Specification of Dependent Variable

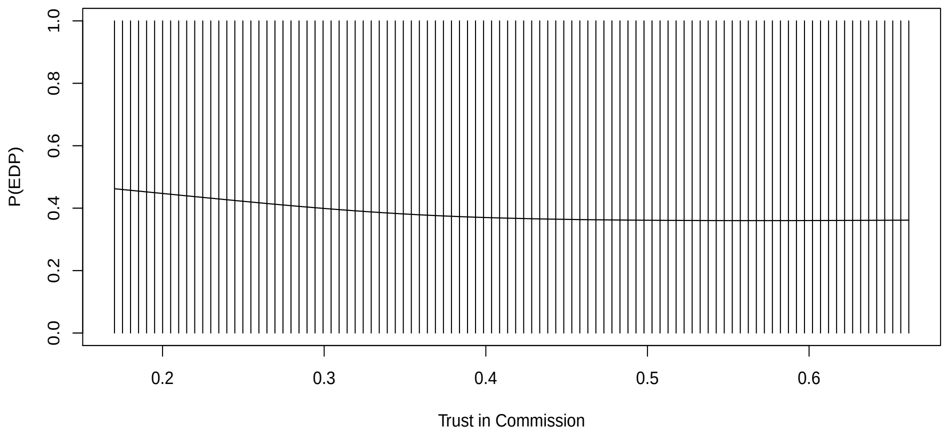


Figure 51: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = high)

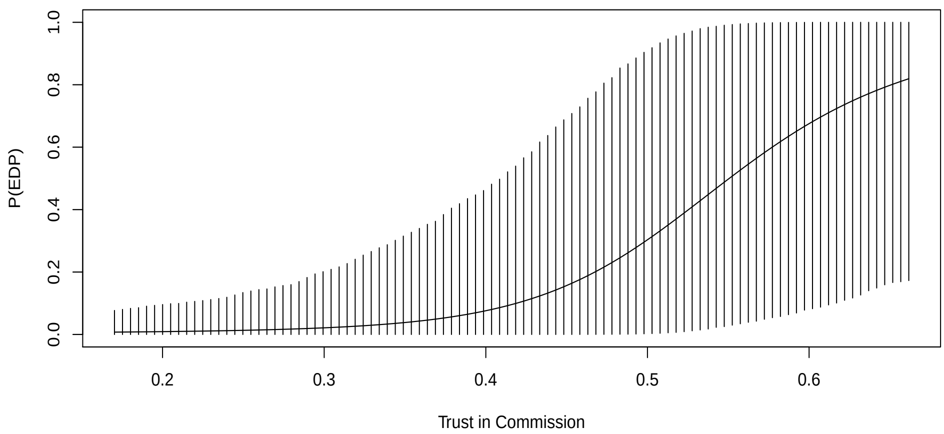


Figure 52: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = low)

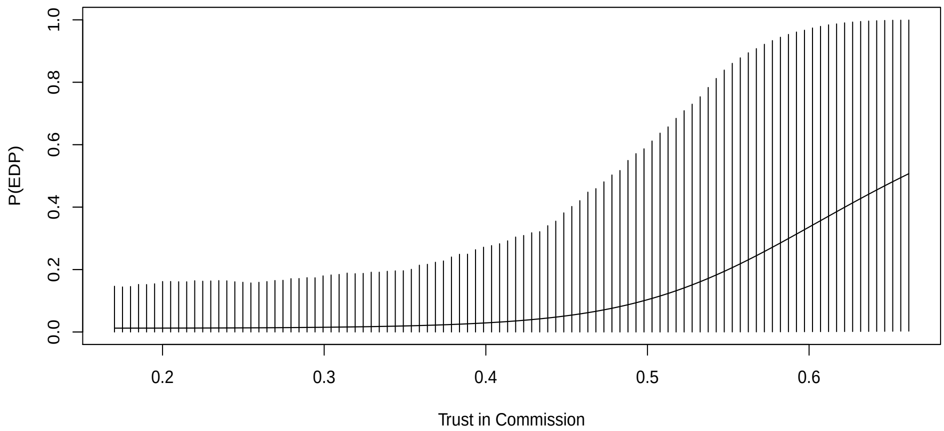


Figure 53: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = high)

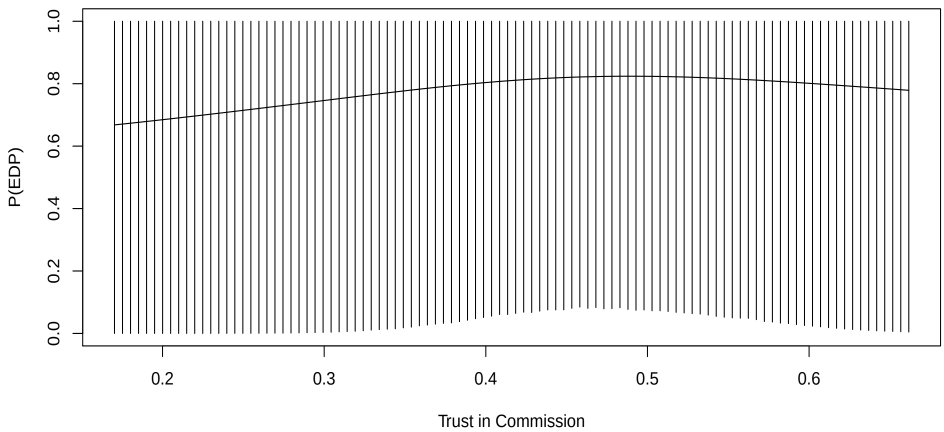


Figure 54: Expected Values for the Effect of the Interaction Between *Trust in European Commission* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = low)

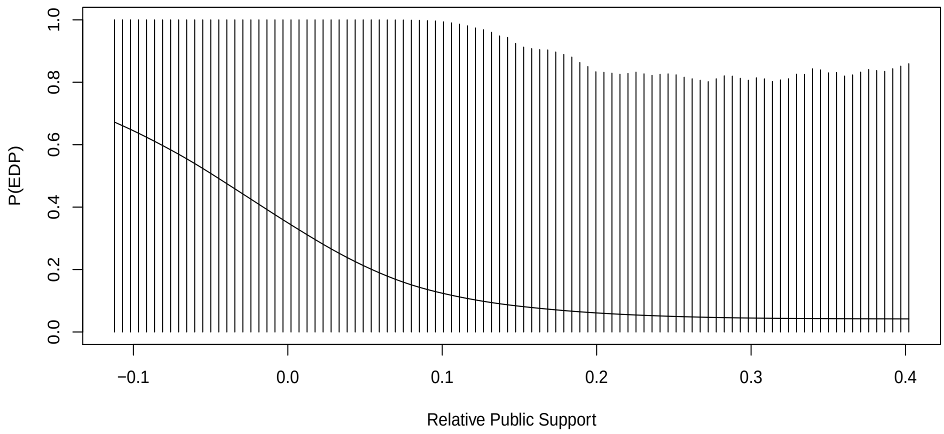


Figure 55: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = high)

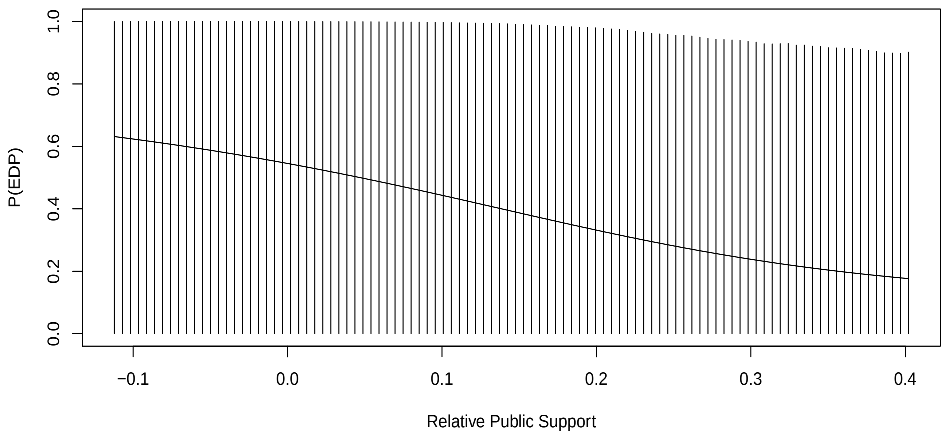


Figure 56: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = low)

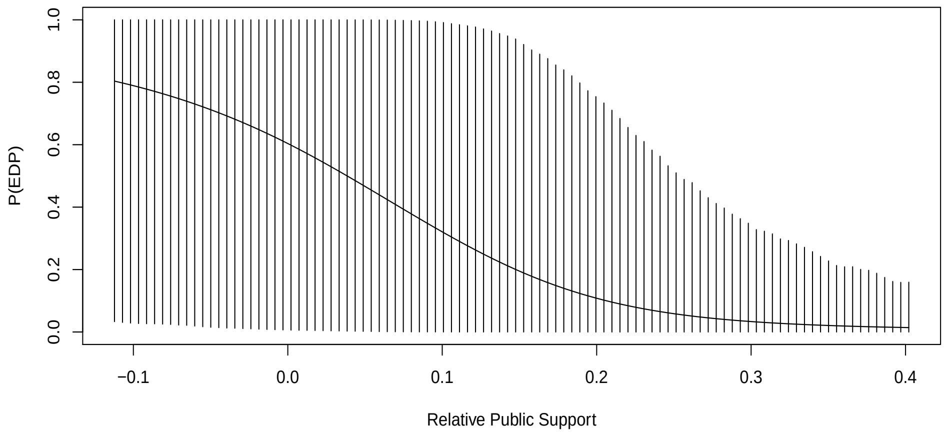


Figure 57: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = high)

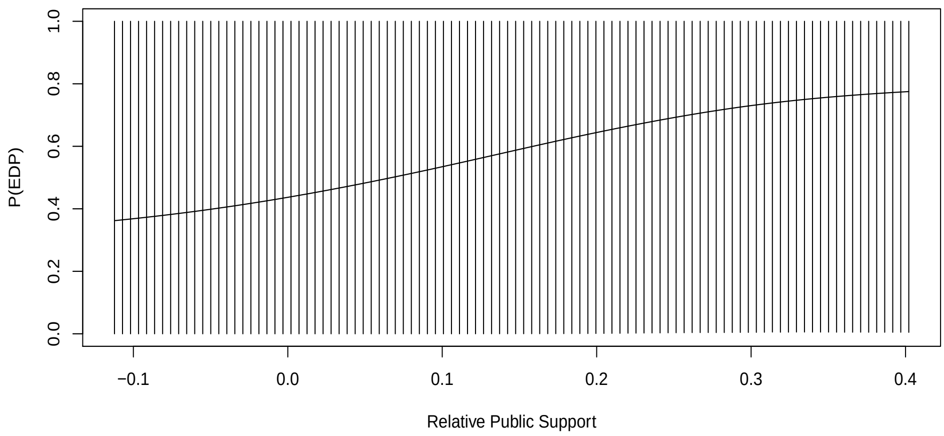


Figure 58: Expected Values for the Effect of the Interaction Between *Relative Public Support* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = low)

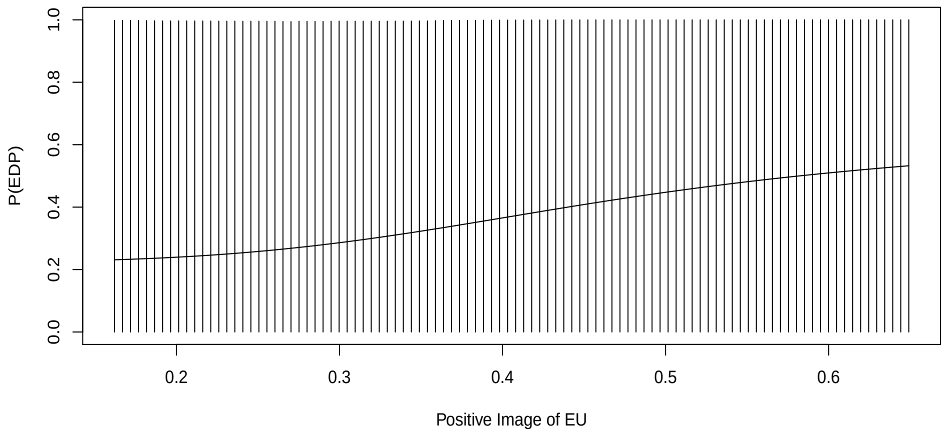


Figure 59: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = high)

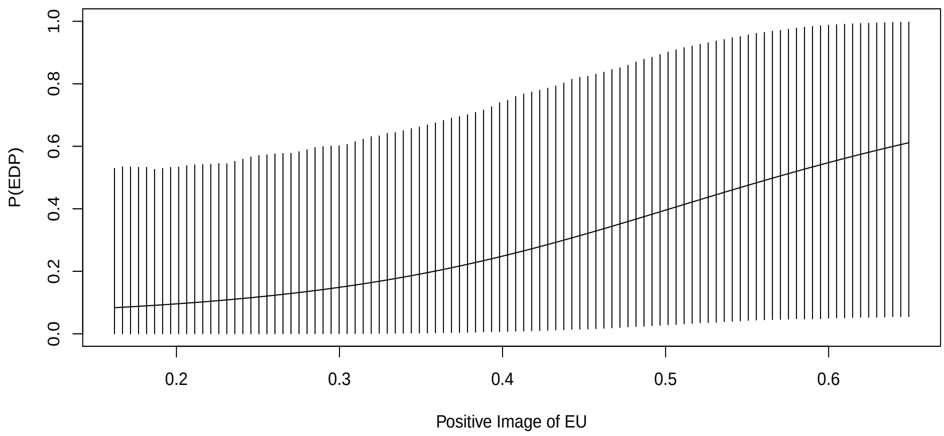


Figure 60: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *Real GDP* on EDP Initiations (Report) (*Real GDP* = low)

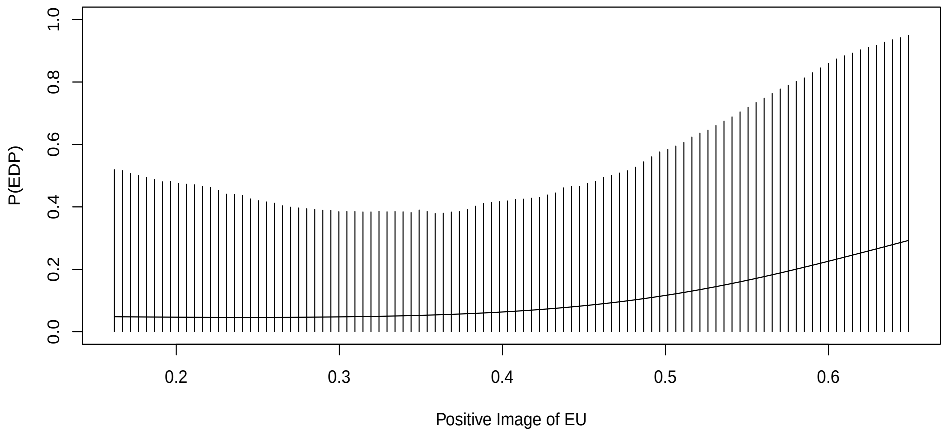


Figure 61: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = high)

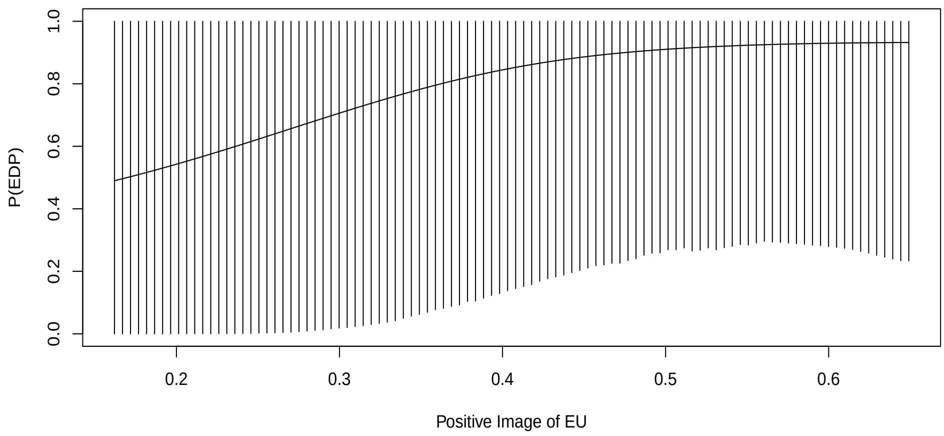
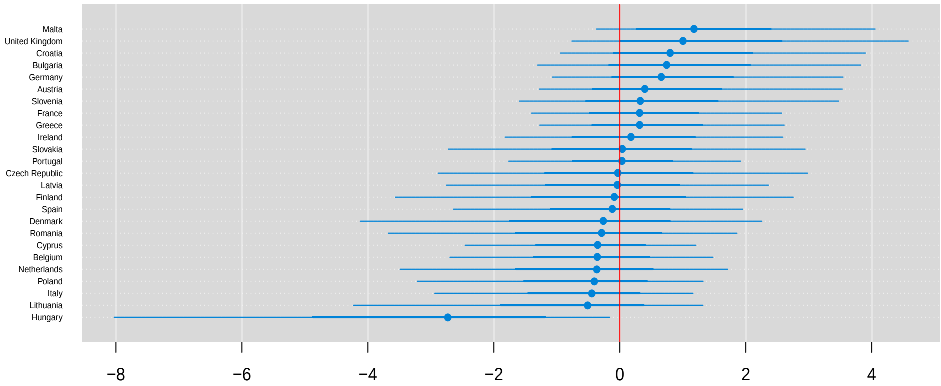


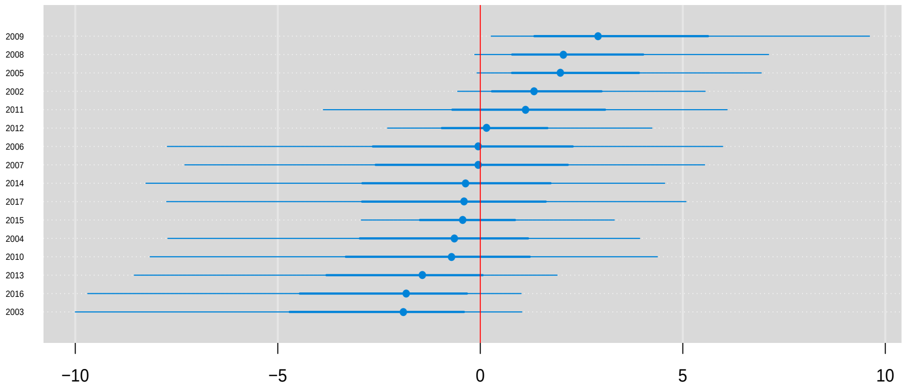
Figure 62: Expected Values for the Effect of the Interaction Between *Positive Image of the EU* and *GDP Growth* on EDP Initiations (Report) (*GDP Growth* = low)

11.3.2 Caterpillar Plots

Alternative Specification of Independent Variables

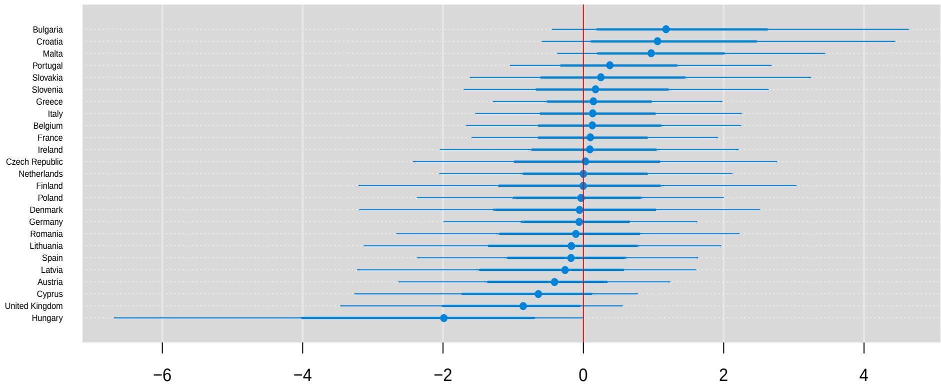


(a) Country-Level Random Effects α_i

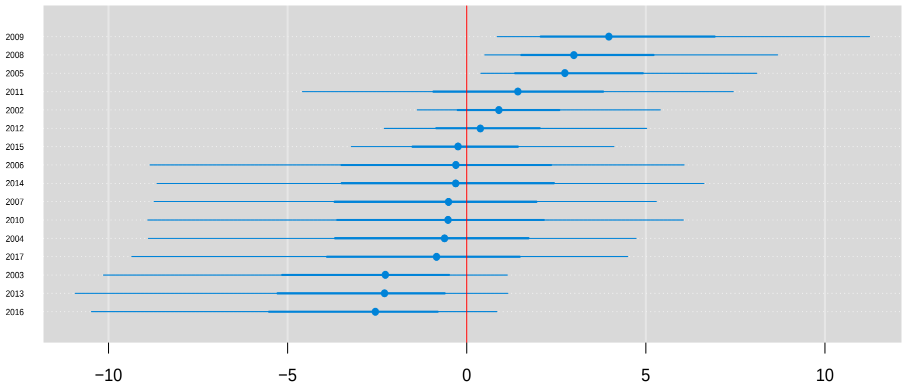


(b) Year-Level Random Effects γ_t

Figure 63: Random Effects of Robustness Check IV (Model 3) (*Trust in European Commission* $\times$ *GDP Growth*).

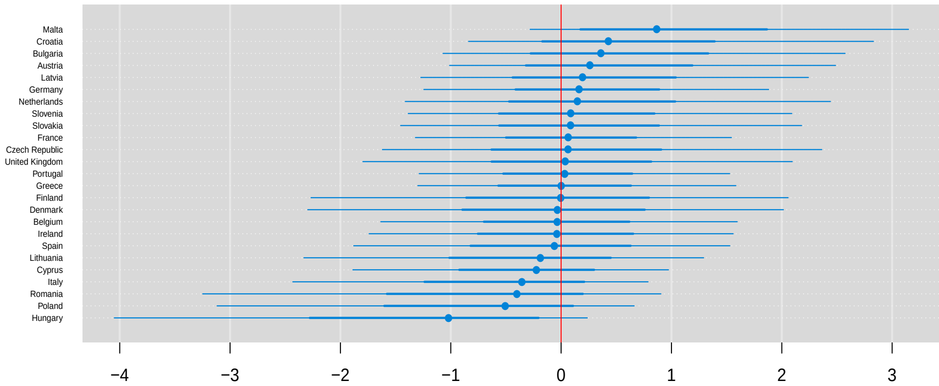


(a) Country-Level Random Effects α_i

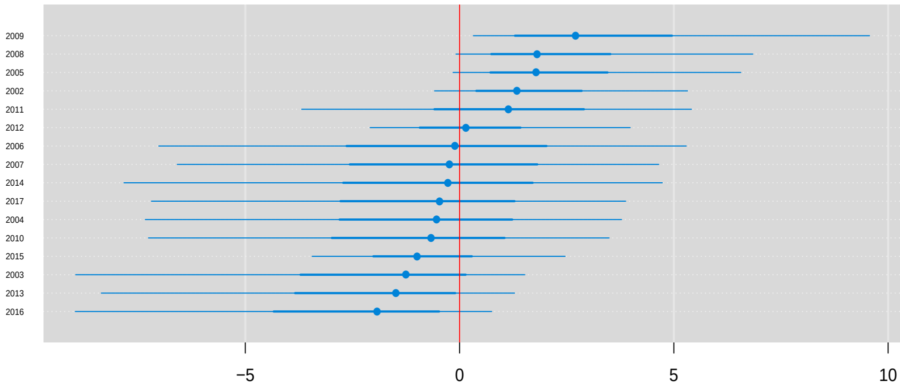


(b) Year-Level Random Effects γ_t

Figure 64: Random Effects of Robustness Check IV (Model 4) (*Relative Public Support* $\times$ *GDP Growth*).

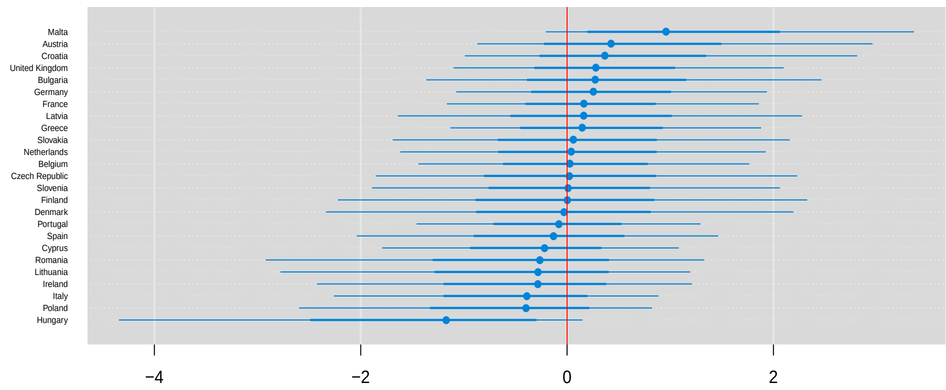


(a) Country-Level Random Effects α_i

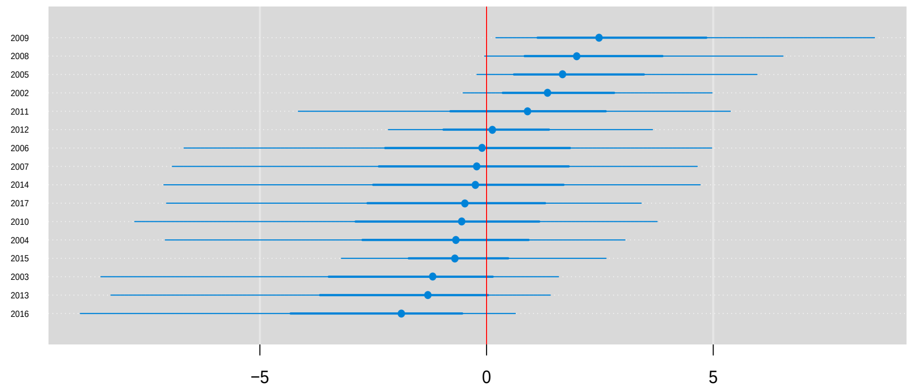


(b) Year-Level Random Effects γ_t

Figure 65: Random Effects of Robustness Check IV (Model 5) (*Positive Image of the EU* $\times$ *Real GDP*).



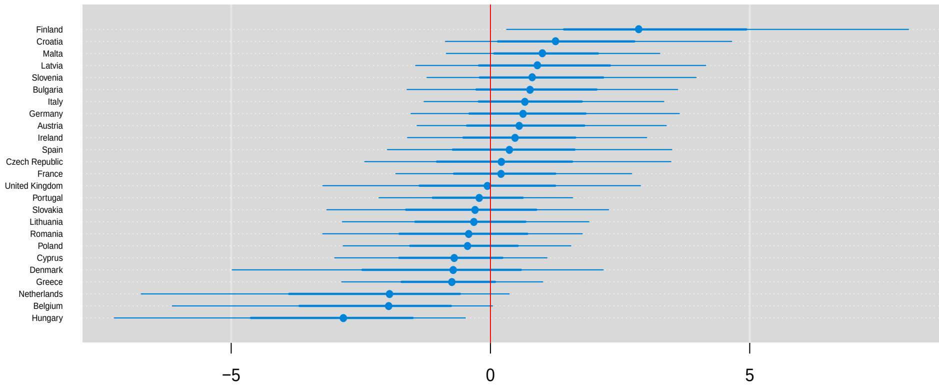
(a) Country-Level Random Effects α_i



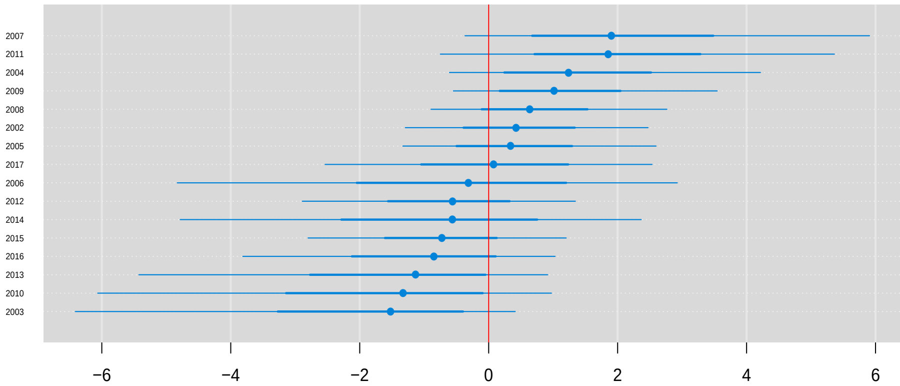
(b) Year-Level Random Effects γ_t

Figure 66: Random Effects of Robustness Check IV (Model 6) (*Positive Image of the EU* $\times$ *GDP Growth*).

Alternative Specification of Dependent Variable

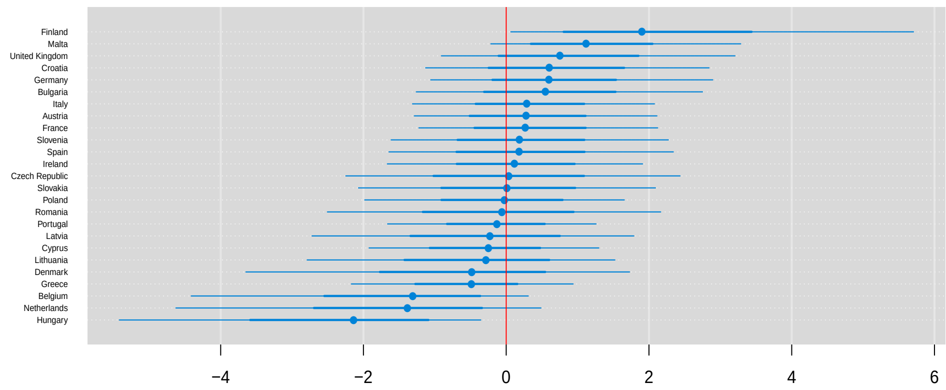


(a) Country-Level Random Effects α_i

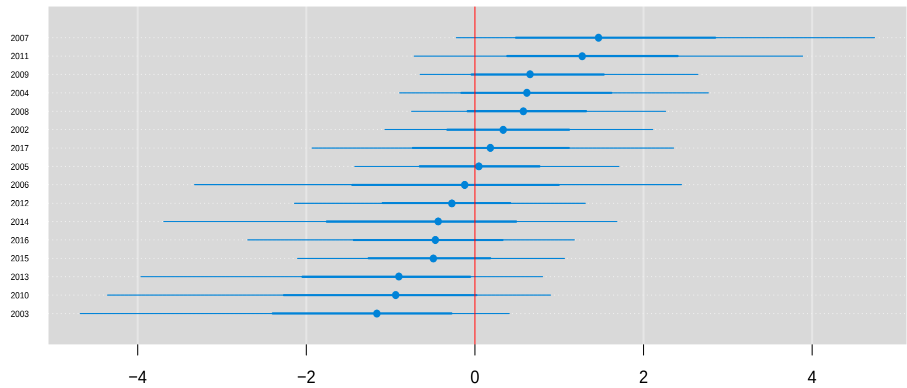


(b) Year-Level Random Effects γ_t

Figure 67: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 7) (*Trust in European Commission* $\times$ *Real GDP*).

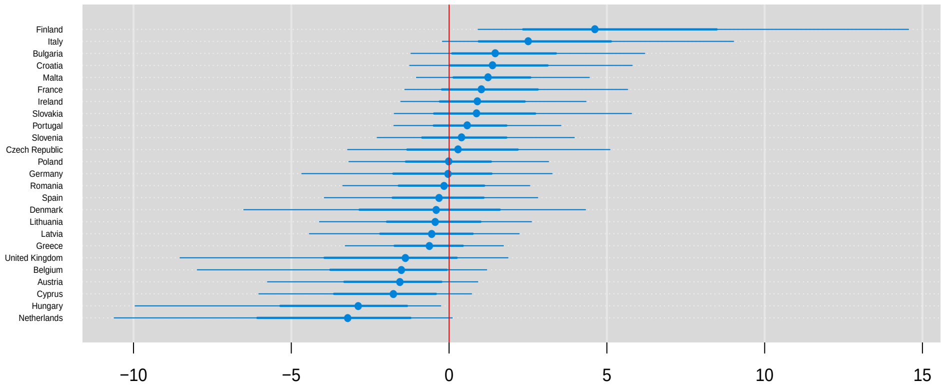


(a) Country-Level Random Effects α_i

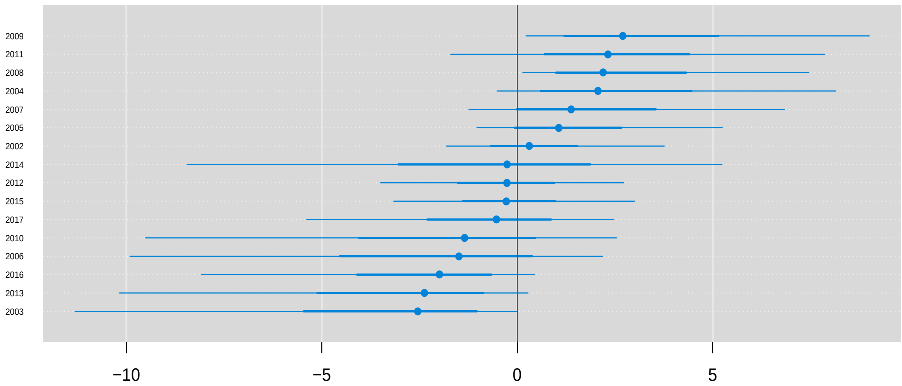


(b) Year-Level Random Effects γ_t

Figure 68: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 8) (*Trust in European Commission* $\times$ *GDP Growth*).

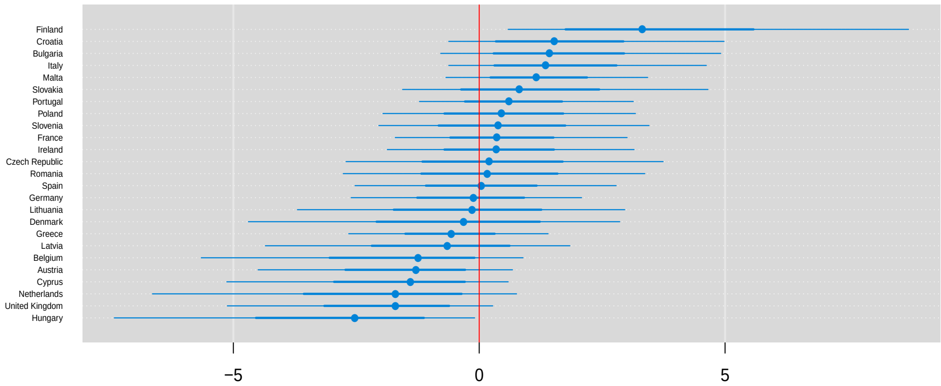


(a) Country-Level Random Effects α_i

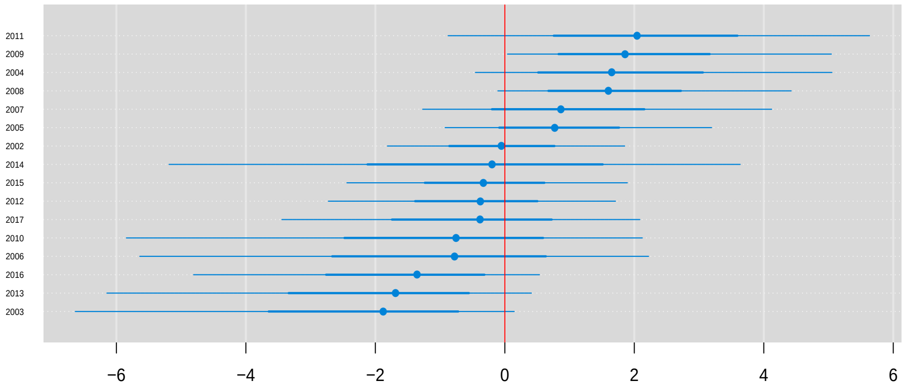


(b) Year-Level Random Effects γ_t

Figure 69: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 9) (*Relative Public Support* $\times$ *Real GDP*).

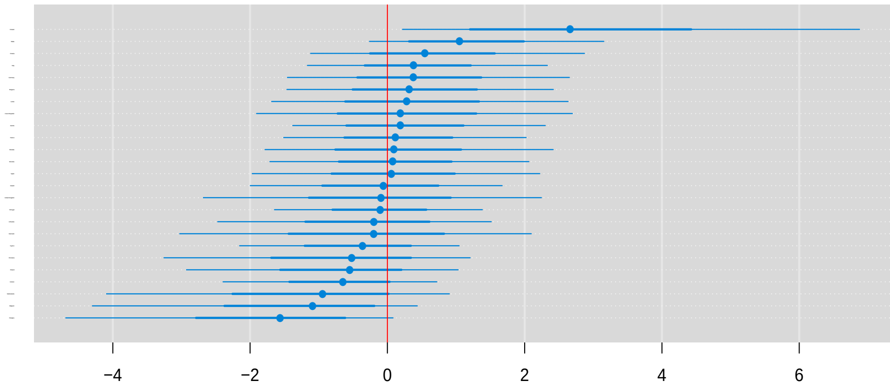


(a) Country-Level Random Effects α_i

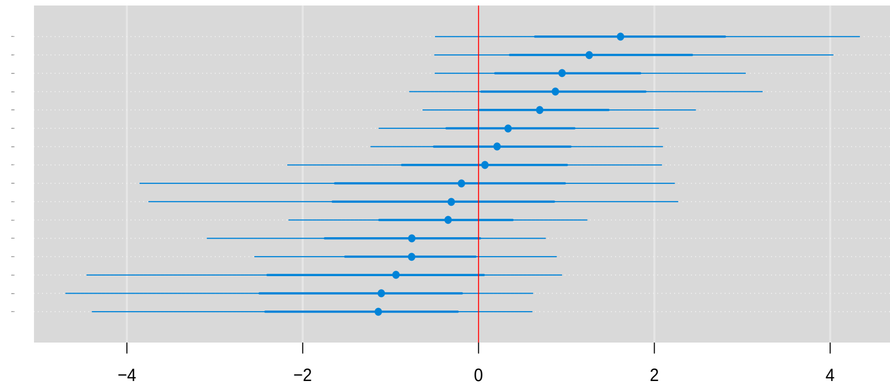


(b) Year-Level Random Effects γ_t

Figure 70: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 10) (*Relative Public Support* $\times$ *GDP Growth*).

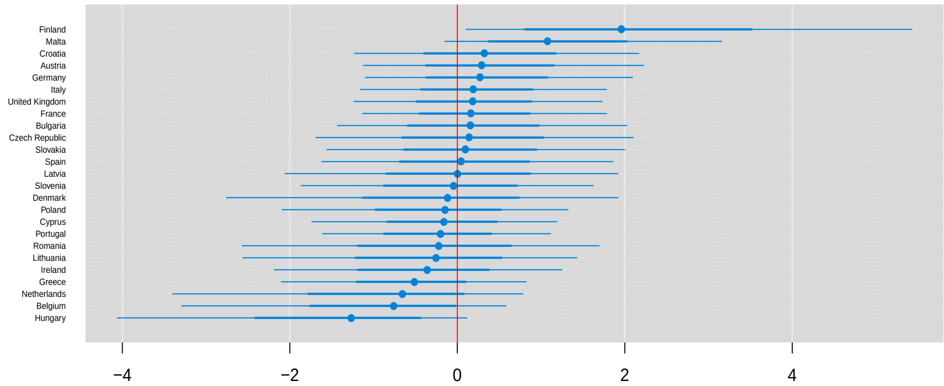


(a) Country-Level Random Effects α_i

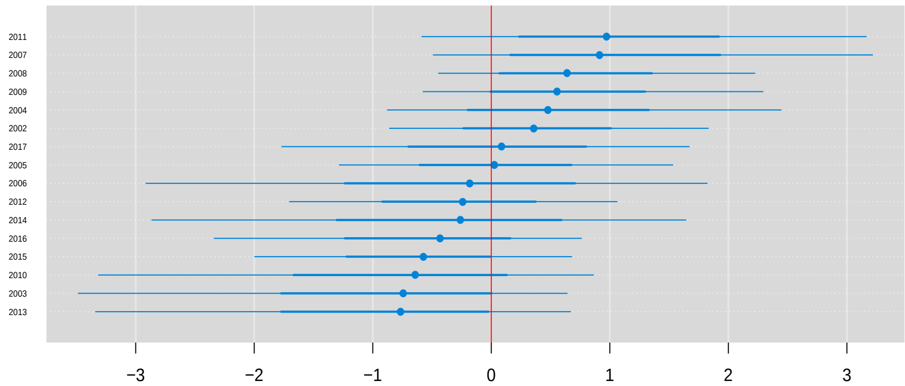


(b) Year-Level Random Effects γ_t

Figure 71: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 11) (*Positive Image of the EU* $\times$ *Real GDP*).



(a) Country-Level Random Effects α_i



(b) Year-Level Random Effects γ_t

Figure 72: Random Effects of Robustness Check DV (Alternative Specification for EDP (Report); Model 12) (*Positive Image of the EU* $\times$ *GDP Growth*).

11.4 Model Diagnostics

11.4.1 Posterior Predictive Checking

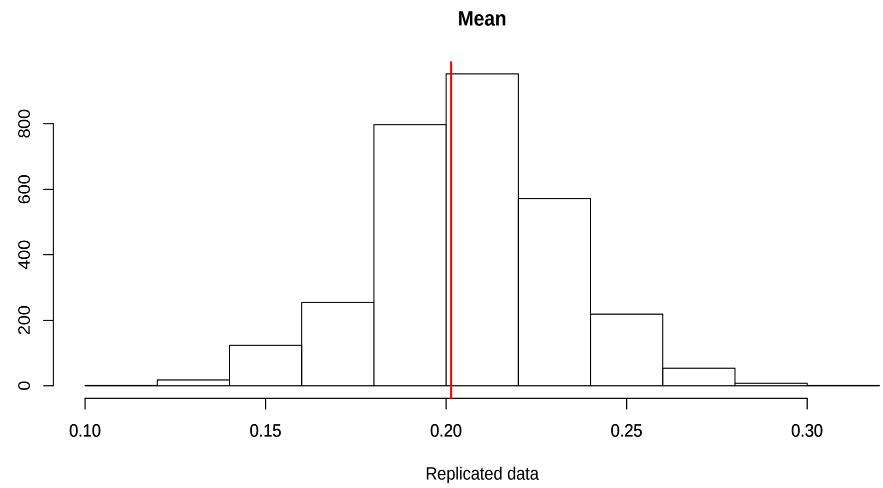


Figure 73: Posterior Predictive Checking for Model 1: Mean

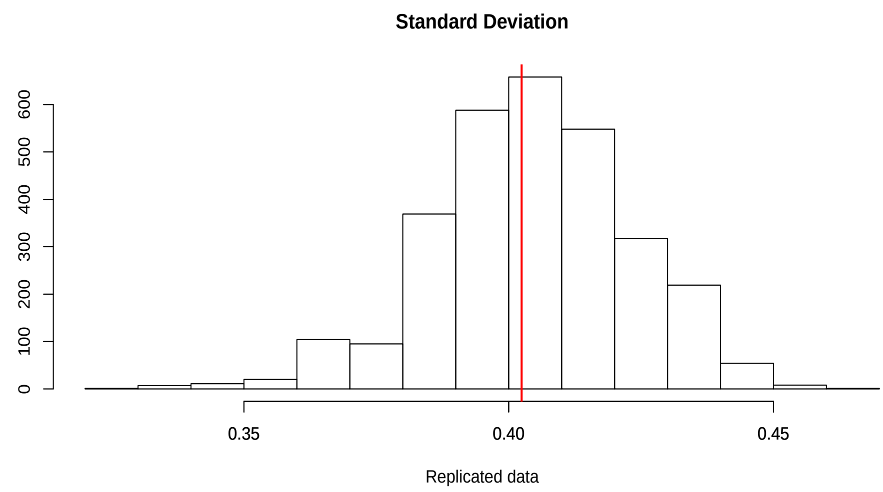


Figure 74: Posterior Predictive Checking for Model 1: Standard Deviation

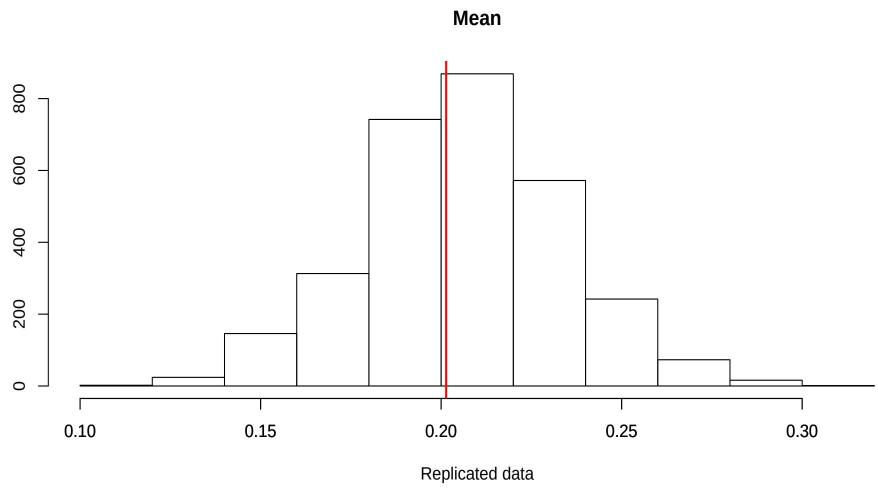


Figure 75: Posterior Predictive Checking for Model 2: Mean

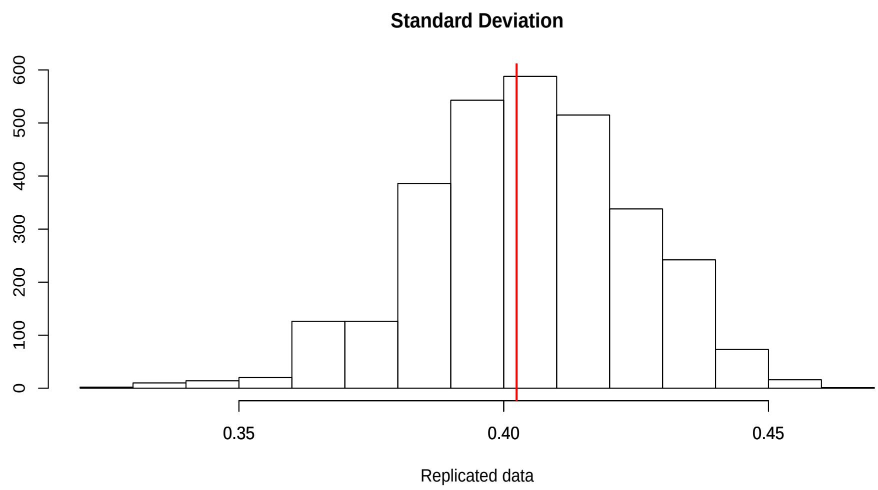


Figure 76: Posterior Predictive Checking for Model 2: Standard Deviation

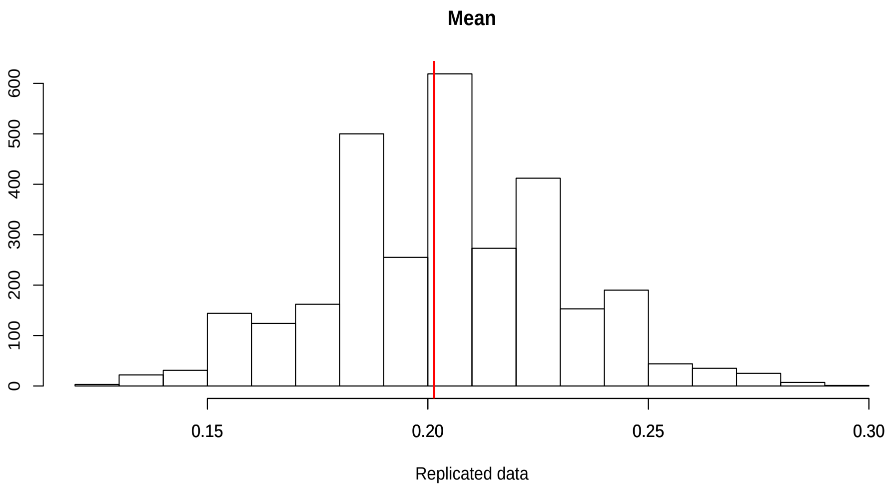


Figure 77: Posterior Predictive Checking for Model 3: Mean

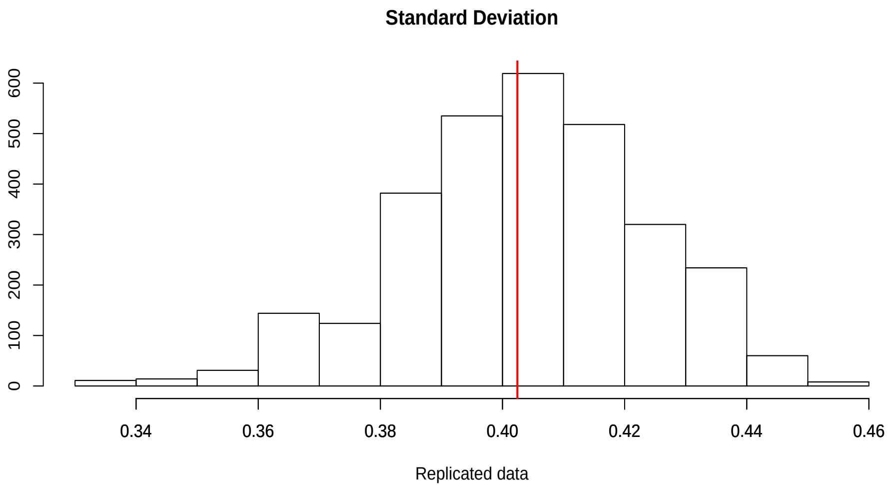


Figure 78: Posterior Predictive Checking for Model 3: Standard Deviation

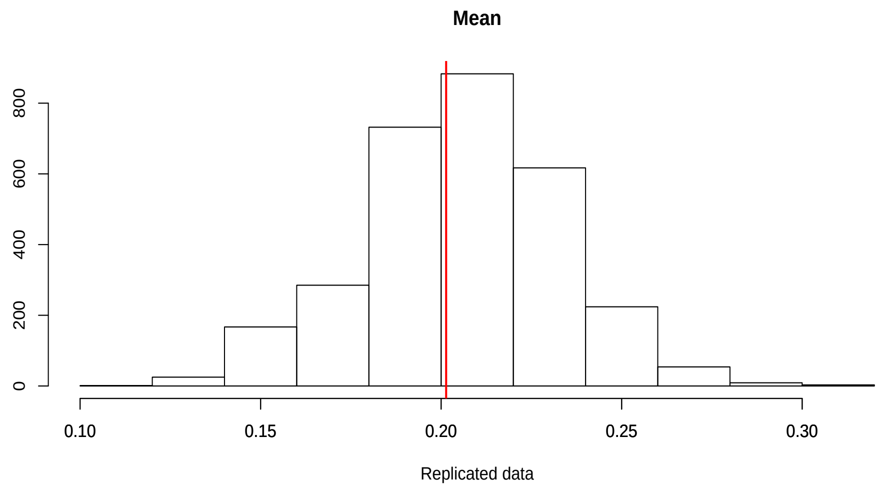


Figure 79: Posterior Predictive Checking for Model 4: Mean

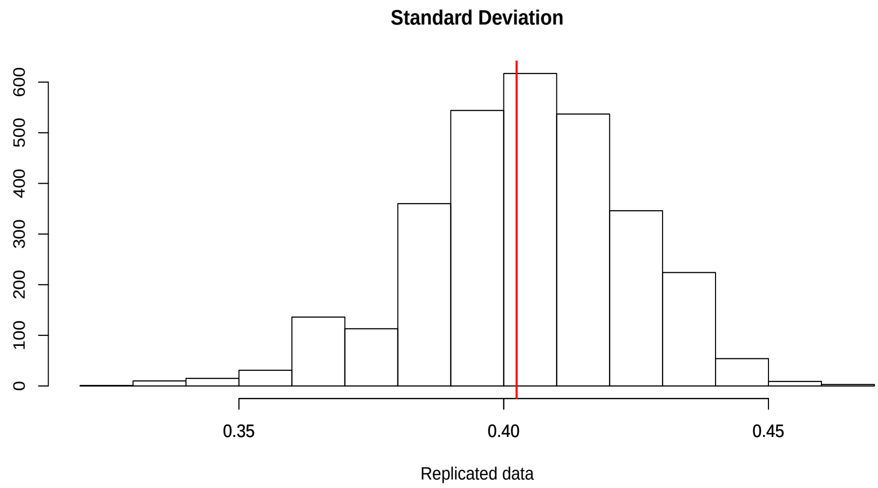


Figure 80: Posterior Predictive Checking for Model 4: Standard Deviation

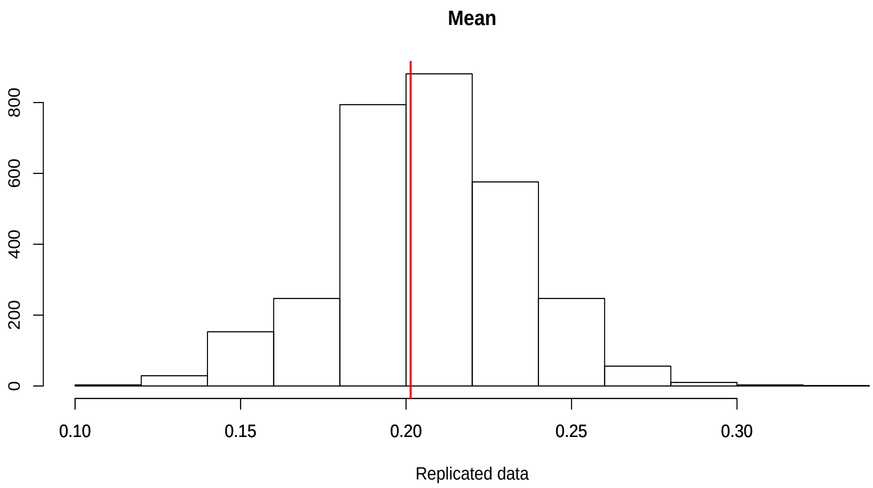


Figure 81: Posterior Predictive Checking for Model 5: Mean

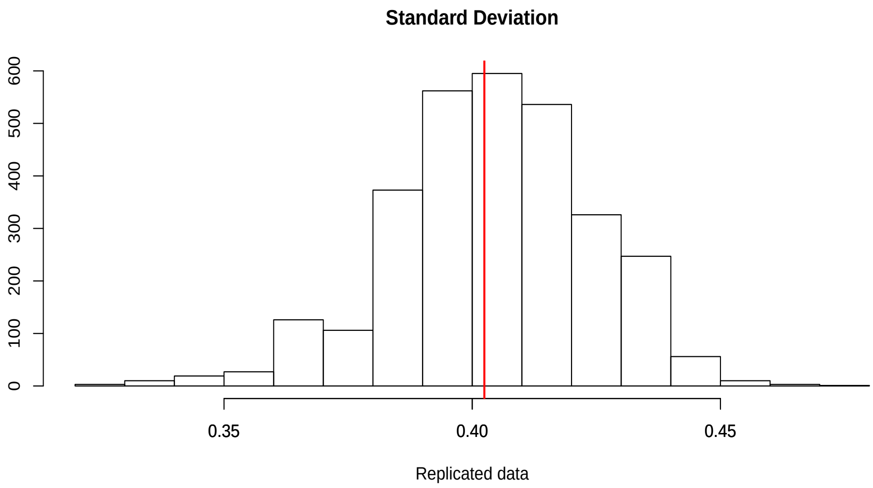


Figure 82: Posterior Predictive Checking for Model 5: Standard Deviation

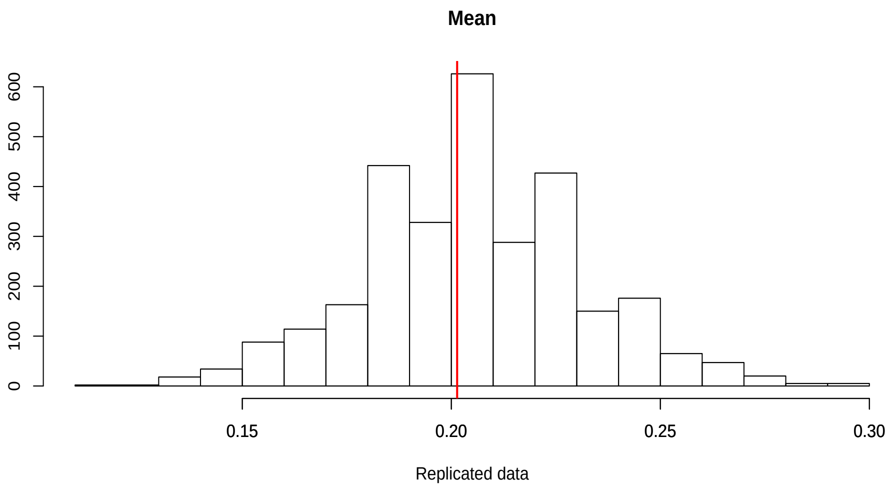


Figure 83: Posterior Predictive Checking for Model 6: Mean

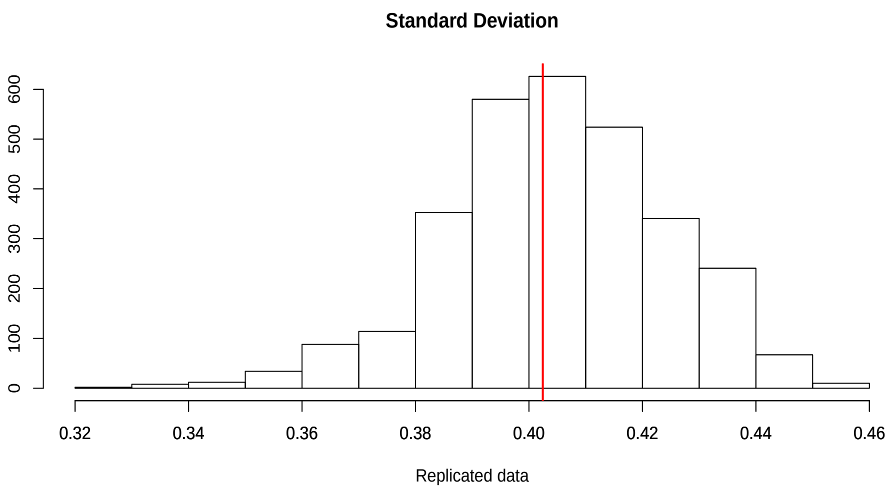


Figure 84: Posterior Predictive Checking for Model 6: Standard Deviation

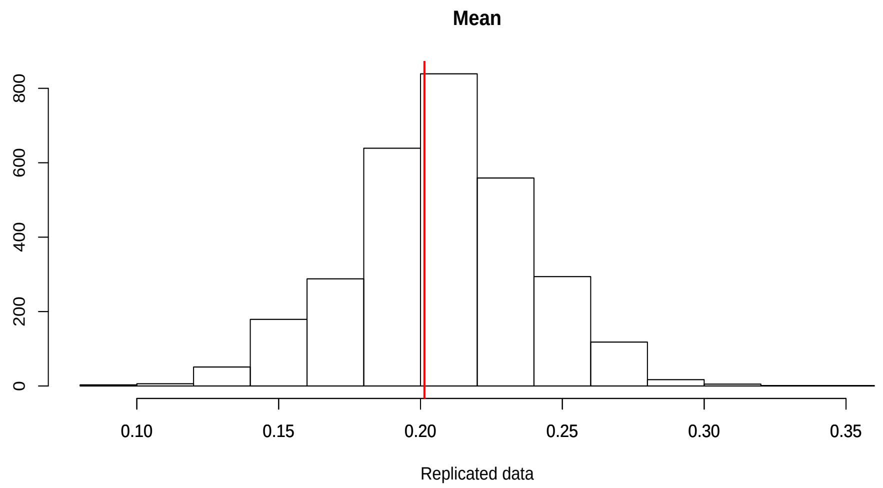


Figure 85: Posterior Predictive Checking for Model 7: Mean

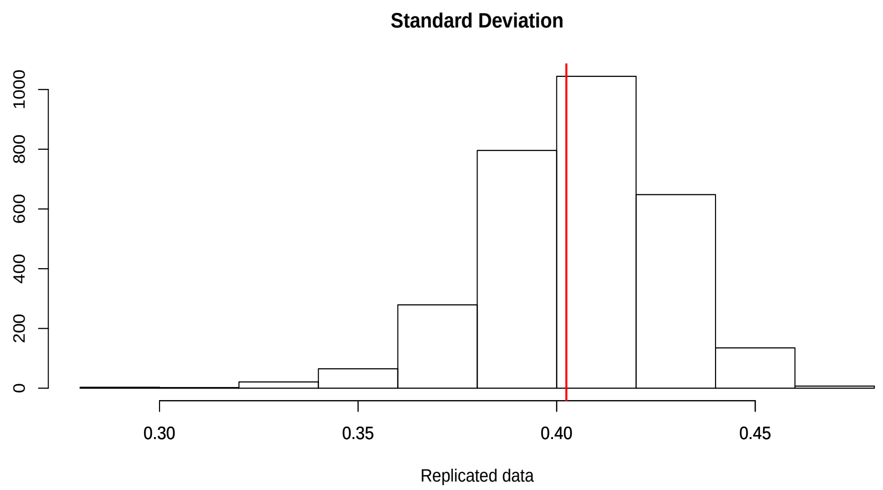


Figure 86: Posterior Predictive Checking for Model 7: Standard Deviation

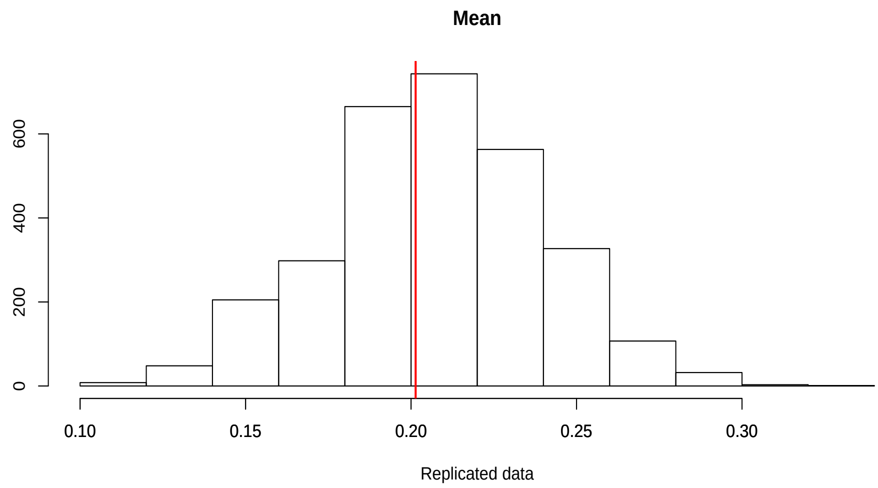


Figure 87: Posterior Predictive Checking for Model 8: Mean

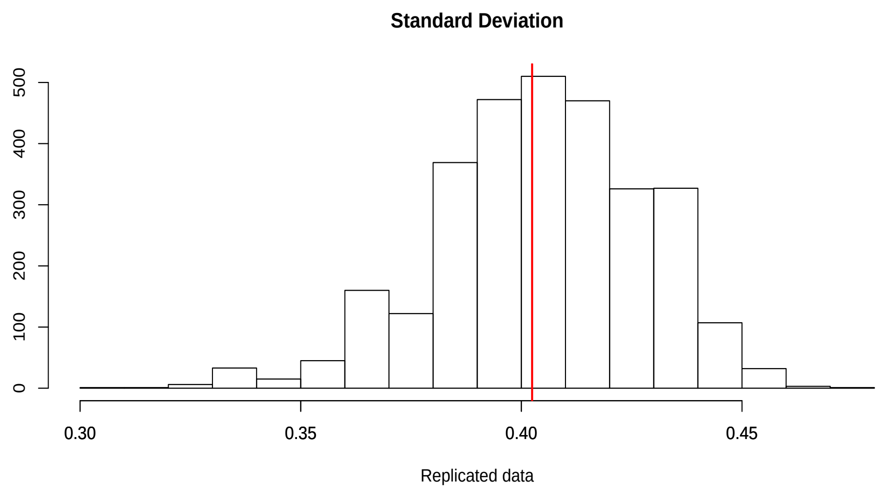


Figure 88: Posterior Predictive Checking for Model 8: Standard Deviation

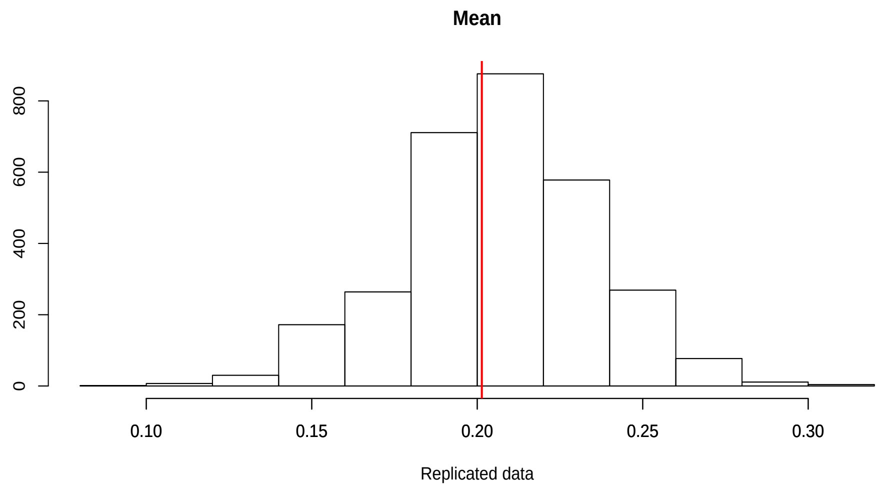


Figure 89: Posterior Predictive Checking for Model 9: Mean

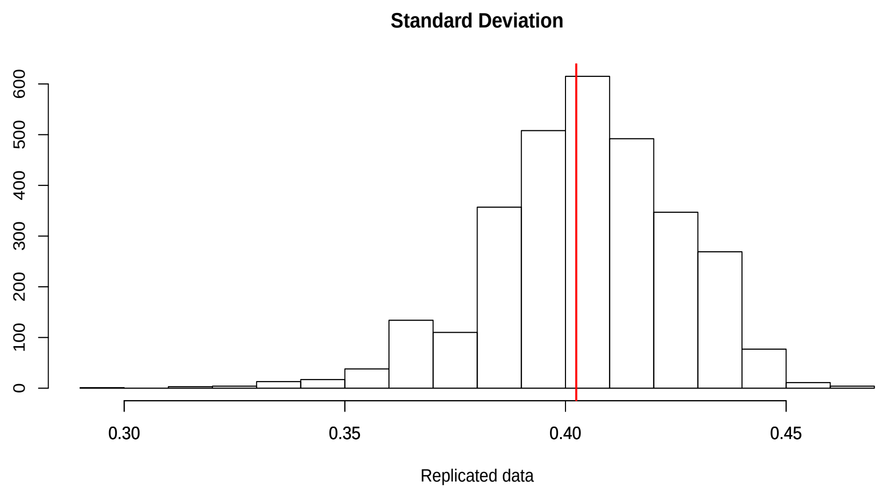


Figure 90: Posterior Predictive Checking for Model 9: Standard Deviation

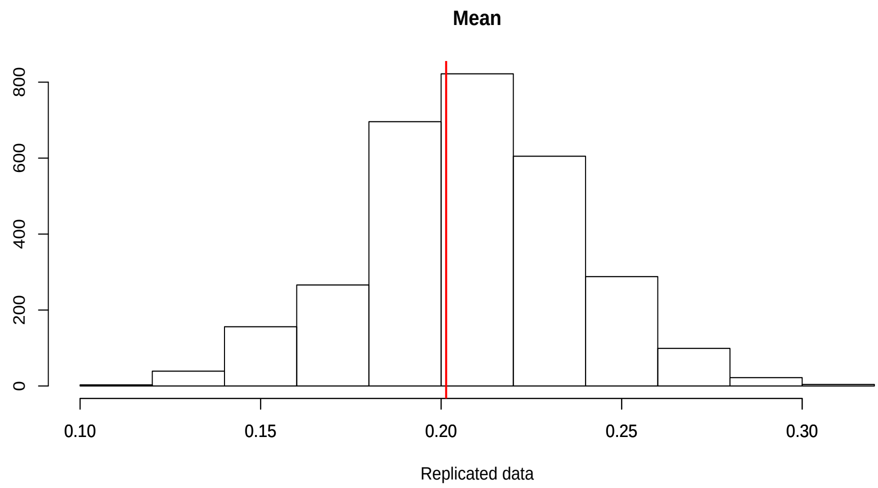


Figure 91: Posterior Predictive Checking for Model 10: Mean

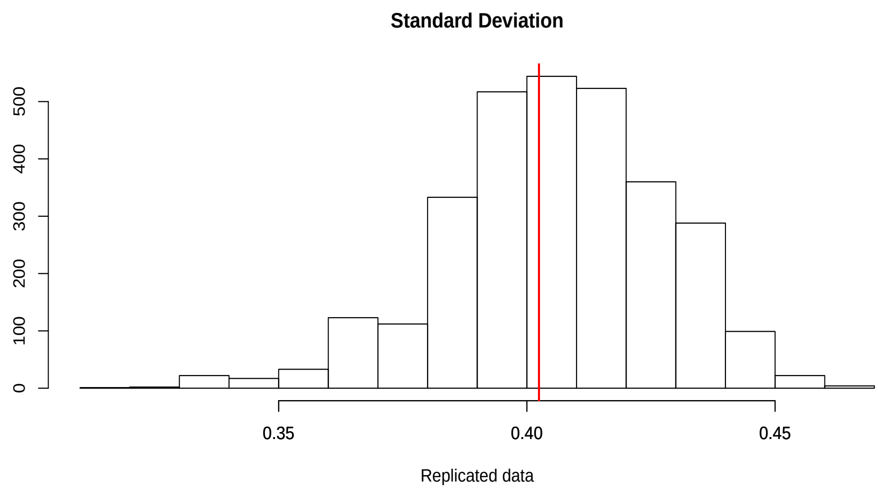


Figure 92: Posterior Predictive Checking for Model 10: Standard Deviation

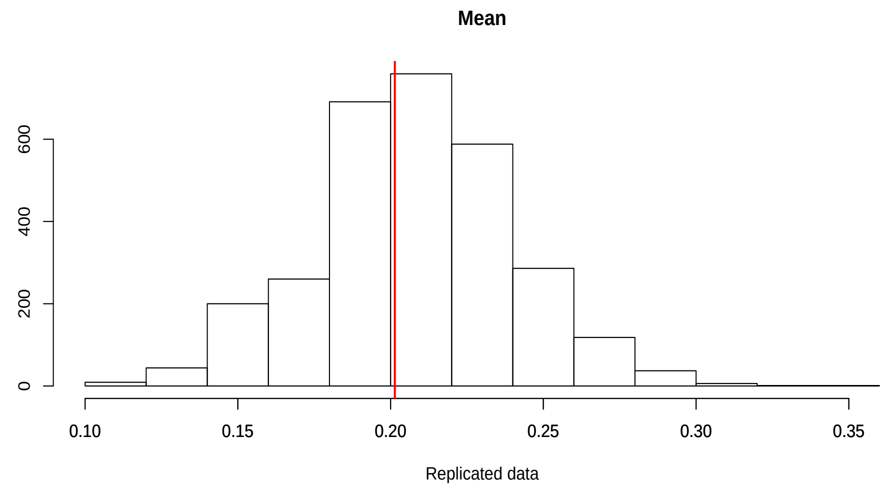


Figure 93: Posterior Predictive Checking for Model 11: Mean

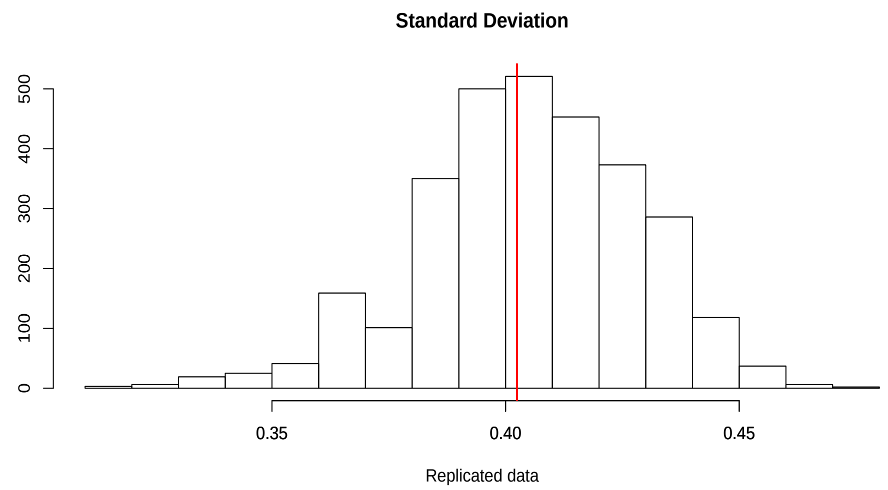


Figure 94: Posterior Predictive Checking for Model 11: Standard Deviation

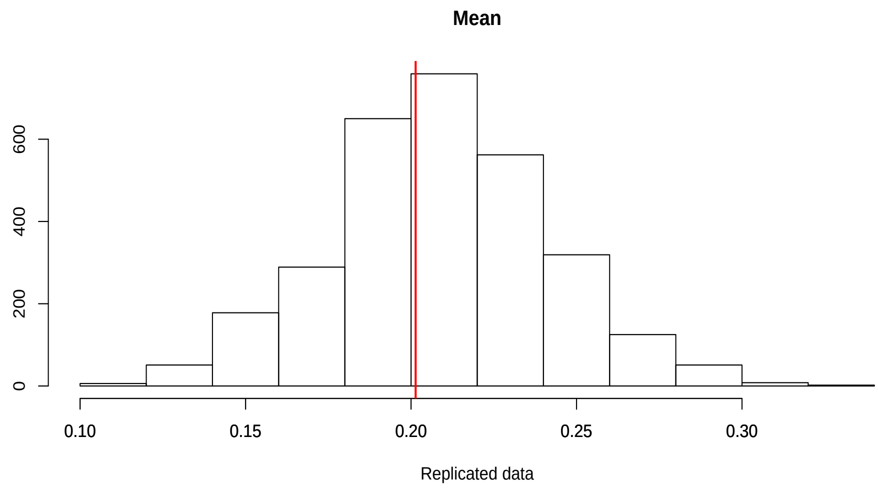


Figure 95: Posterior Predictive Checking for Model 12: Mean

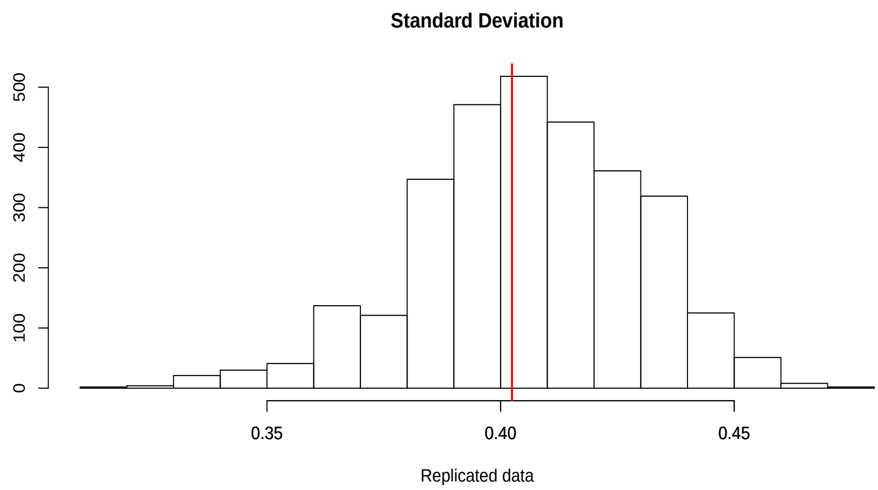


Figure 96: Posterior Predictive Checking for Model 12: Standard Deviation

Table 22: Bayesian p-Values

Model	Mean	Standard Deviation
1	0.60	0.60
2	0.59	0.59
3	0.59	0.59
4	0.59	0.60
5	0.59	0.59
6	0.60	0.60
7	0.61	0.61
8	0.59	0.59
9	0.60	0.60
10	0.61	0.61
11	0.60	0.60
12	0.61	0.61

11.4.2 Deviance Information Criterion

Table 23: Deviance Information Criterion

Model	DIC
1	99.17
2	102.05
3	101.63
4	103.04
5	98.19
6	97.29
7	130.14
8	125.18
9	147.87
10	129.36
11	136.50
12	131.54

Bibliography

Ai, C. and E. C. Norton (2003). "Interaction Terms in Logit and Probit Models". *Economics Letters 80*(1), 123–129.

Alesina, A., G. D. Cohen, and N. Roubini (1993). "Electoral Business Cycle in Industrial Democracies". *European Journal of Political Economy 9*(1), 1–23.

Alesina, A., R. Hausmann, R. Hommes, and E. Stein (1999). "Budget Institutions and Fiscal Performance in Latin America". *Journal of Development Economics 59*(2), 253–273.

Alesina, A. and G. Tabellini (1990). "A Positive Theory of Fiscal Deficits and Government Debt". *The Review of Economic Studies 57*(3), 403–414.

Alesina, A. F. and R. Perotti (1995). "Fiscal Expansions and Fiscal Adjustments in OECD Countries". *NBER Working Paper* w5214, Cambridge, MA: National Bureau of Economic Research.

Alt, J. E. and D. D. Lassen (2006). "Transparency, Political Polarization, and Political Budget Cycles in OECD Countries". *American Journal of Political Science 50*(3), 530–550.

Annett, A. (2001). "Social fractionalization, political instability, and the size of government". *IMF Staff papers* 3, Washington, D.C.: International Monetary Fund.

Annett, A. (2006). "Enforcement and the Stability and Growth Pact: How Fiscal Policy Did and Did Not Change Under Europe's Fiscal Framework". *IMF Working Paper* WP/06/116, Washington, D.C.: International Monetary Fund.

Aquino, J., D. Enzmann, M. Schwartz, N. Jain, and S. Kraft (2016). *"descr: Descriptive Statistics"*. R package version 1.1.3.

Arel-Bundock, V. (2017). *"countrycode: Convert Country Names and Country Codes"*. R package version 0.19.

Bäck, H. and P. Dumont (2007). "Combining Large-N and Small-N Strategies: The Way Forward in Coalition Research". *West European Politics 30*(3), 467–501.

Baerg, N. R. and M. Hallerberg (2016). "Explaining Instability in the Stability and Growth Pact: the Contribution of Member State Power and Euroskepticism to the Euro Crisis". *Comparative Political Studies 49*(7), 968–1009.

Baron, R. M. and D. A. Kenny (1986). "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations". *Journal of Personality and Social Psychology 51*(6), 1173–1182.

Barro, R. J. (1979). "On the Determination of the Public Debt". *Journal of Political Economy 87*(5), 940–971.

Baskaran, T. and Z. Hessami (2017). "Monetary Integration, Soft Budget Constraints, and the EMU Sovereign Debt Crises". *Economics & Politics 29*(3), 252–275.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software 67*(1), 1–48.

Bauer, M. W. and S. Becker (2014). "The Unexpected Winner of the Crisis: The European Commission's Strengthened Role in Economic Governance". *Journal of European Integration 36*(3), 213–229.

Beaujean, A. A. (2012). *"BaylorEdPsych: R Package for Baylor University Educational Psychology Quantitative Courses"*. R package version 0.5.

Beck, N. (2001a). "Time-Series–Cross-Section Data". *Statistica Neerlandica 55*(2), 111–133.

Beck, N. (2001b). "Time-Series–Cross-Section Data: What Have We Learned in the Past Few Years?". *Annual Review of Political Science 4*(1), 271–293.

Beck, N. and J. N. Katz (1995). "What to Do (and Not to Do) with Time-Series Cross-Section Data". *American Political Science Review 89*(3), 634–647.

Beck, N. and J. N. Katz (2007). "Random Coefficient Models for Time-Series–Cross-Section Data: Monte Carlo Experiments". *Political Analysis 15*(2), 182–195.

Beck, N., J. N. Katz, and R. Tucker (1998). "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable". *American Journal of Political Science 42*(4), 1260–1288.

Blanchard, O. (1993). "Consumption and the Recession of 1990-1991". *The American Economic Review 83*(2), 270–274.

Börzel, T. A. (2001). "Non-Compliance in the European Union: Pathology or Statistical Artefact?". *Journal of European Public Policy 8*(5), 803–824.

Börzel, T. A. (2003a). "Guarding the Treaty: the Compliance Strategies of the European Commission". In T. A. Börzel and R. A. Cichowski (Eds.), *The State of the European Union. Law, Politics and Society.*, Volume 6, pp. 197–220. Oxford: Oxford University Press.

Börzel, T. A. (2003b). "How the European Union Interacts With Its Member States". Research Paper, Institute for Advanced Studies, Vienna. `https://nbn-resolving.org/urn:nbn:de:0168-ssoar-246018`, accessed on December 11th, 2018.

Bova, E., T. Kinda, P. Muthoora, and F. Toscani (2015). "Fiscal Rules at a Glance". *Background document updating IMF Working Paper 12/273*, Washington, D.C.: International Monetary Fund.

Brambor, T., W. R. Clark, and M. Golder (2006). "Understanding Interaction Models: Improving Empirical Analyses". *Political Analysis 14*(1), 63–82.

Braun, D. and M. Tausendpfund (2014). "The Impact of the Euro Crisis on Citizens' Support for the European Union". *Journal of European Integration 36*(3), 231–245.

Breiman, L. (2001). "Random Forests". *Machine Learning 45*(1), 5–32.

Buchanan, J. M. and R. E. Wagner (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes.* New York, NY: Academic Press.

Buti, M. and N. Carnot (2012). "The EMU Debt Crisis: Early Lessons and Reforms". *Journal of Common Market Studies 50*(6), 899–911.

Buti, M., D. Franco, and H. Ongena (1998). "Fiscal Discipline and Flexibility in EMU: the Implementation of the Stability and Growth Pact". *Oxford Review of Economic Policy 14*(3), 81–97.

Buti, M. and G. Giudice (2002). "Maastricht's Fiscal Rules at Ten: An Assessment". *Journal of Common Market Studies 40*(5), 110–124.

Caldeira, G. A. and J. L. Gibson (1992). "The Etiology of Public Support for the Supreme Court". *American Journal of Political Science 36*(3), 635–664.

Carsey, T. M. and J. J. Harden (2013). *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks, CA: Sage Publications.

Carter, D. B. and C. S. Signorino (2010). "Back to the Future: Modeling Time Dependence in Binary Data". *Political Analysis 18*(3), 271–292.

Christiansen, T. (1997). "Tensions of European Governance: Politicized Bureaucracy and Multiple Accountability in the European Commission". *Journal of European Public Policy 4*(1), 73–90.

Clements, B., K. Nanou, and S. Verney (2014). "We No Longer Love You, But We Don't Want to Leave You': the Eurozone Crisis and Popular Euroscepticism in Greece". *Journal of European Integration 36*(3), 247–265.

Collins, L. M., J. L. Schafer, and C.-M. Kam (2001). "A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures". *Psychological Methods 6*(4), 330.

Cowles, M. K. and B. P. Carlin (1996). "Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review". *Journal of the American Statistical Association 91*(434), 883–904.

Cukierman, A. and A. H. Meltzer (1989). "A Political Theory of Government Debt and Deficits in a Neo-Ricardian Framework". *The American Economic Review 79*(4), 713–732.

Curtis, S. M. (2015). *"mcmcplots: Create Plots from MCMC Output"*. R package version 0.4.2.

Dahl, D. B. (2016). *"xtable: Export Tables to LaTeX or HTML"*. R package version 1.8-2.

De Haan, J., H. Berger, and D.-J. Jansen (2004). "Why has the Stability and Growth Pact Failed?". *International Finance 7*(2), 235–260.

De Haan, J. and J.-E. Sturm (1994). "Political and institutional determinants of fiscal policy in the European Community. *Public Choice 80*(1-2), 157–172.

De Leeuw, J. and E. Meijer (2008). "Introduction to Multilevel Analysis". In J. D. Leeuw and E. Meijer (Eds.), *Handbook of Multilevel Analysis*, pp. 1–75. Heidelberg: Springer.

De Vries, C. E. and K. V. Kersbergen (2007). "Interests, Identity and Political Allegiance in the European Union". *Acta Politica 42*(2-3), 307–328.

Dixon, P. M. (2002). "Bootstrap Resampling". In A. H. El-Shaarawi and W. W. Piegorsch (Eds.), *Encyclopedia of Environmetrics*, pp. 1–9. Hoboken, NJ: John Wiley & Sons, Ltd.

Doove, L. L., S. van Buuren, and E. Dusseldorp (2014). "Recursive Partitioning for Missing Data Imputation in the Presence of Interaction Effects". *Computational Statistics & Data Analysis 72*(2014), 92–104.

Eichengreen, B. (2005). "Europe, the Euro and the ECB: Monetary Success, Fiscal Failure". *Journal of Policy Modeling 27*(4), 427–439.

Eichengreen, B. and C. Wyplosz (1998). "the stability pact: More than a minor nuisance?". *Economic Policy 13*(26), 66–113.

Eijffinger, S. C. and J. De Haan (1996). "The Political Economy of Central-Bank Independence". *Special Papers in International Economics* No. 19, Princeton, NJ: International Finance Section, Department of Economics, Princeton University.

European Central Bank (2003). *ECB Monthly Bulletin June 2003*. `https://www.ecb.europa.eu/pub/pdf/other/mb200306_focus07.en.pdf`, accessed on December 11th, 2018.

European Commission (2017a). *Assessment of Effective Action*. `https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact/corrective-arm-excessive-deficit-procedure/assessment-effective-action_en`, accessed on December 11th, 2018.

European Commission (2017b). *Excessive Deficit Procedures - Overview.* https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact/corrective-arm-excessive-deficit-procedure/excessive-deficit-procedures-overview_en, accessed on December 11th, 2018.

European Commission (2017c). *Public Opinion. Eurobarometer Interactive.* http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/Chart/index, accessed on December 11th, 2018.

European Commission (2017d). *Stability and Growth Pact.* https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact_en, accessed on December 11th, 2018.

European Commission (2018a). *Excessive Deficit Procedure Under the Two Pack Regulation.* https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact/corrective-arm-excessive-deficit-procedure/excessive-deficit-procedure-under-two-pack-regulation_en, accessed on August 14th, 2018.

European Commission (2018b). *Public Opinion. Interactive Eurobarometer. "And, for each of them, please tell me if you tend to trust it or tend not to trust it?" The European Commission.* http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Chart/getChart/chartType/gridChart//themeKy/9/groupKy/24/savFile/661, accessed on December 11th, 2018.

European Commission (2018c). *Public Opinion. Interactive Eurobarometer. "In general, does the European Union conjure up for you a very positive, fairly positive, neutral, fairly negative or very negative image?".* http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Chart/getChart/chartType/gridChart//themeKy/19/groupKy/102/savFile/196, accessed on December 11th, 2018.

European Union (2008). *Consolidated Version of the Treaty on the Functioning of the European Union - PART THREE: UNION POLICIES AND INTERNAL ACTIONS - TITLE VIII: ECONOMIC AND MONETARY POLICY - Chapter 1: Economic Policy*

- *Article 126 (Ex Article 104 TEC).* http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008E126&from=EN, accessed on December 11th, 2018.

European Union (2018a). *The EU in Brief*. https://europa.eu/european-union/about-eu/eu-in-brief_en, accessed on December 11th, 2018.

European Union (2018b). *What is the Euro Area?* https://ec.europa.eu/info/business-economy-euro/euro-area/what-euro-area_en, accessed on December 11th, 2018.

Eurostat (2018). *General Government Gross Debt (EDP Concept), Consolidated - Annual Data.* http://ec.europa.eu/eurostat/web/products-datasets/-/tipsgo10, accessed on December 11th, 2018.

Fabbrini, S. (2013). "Intergovernmentalism and its Limits: Assessing the European Union's Answer to the Euro Crisis". *Comparative Political Studies 46*(9), 1003–1029.

Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). *Regression: Models, Methods and Applications*. Heidelberg: Springer.

Faini, R. (2006). "Fiscal Policy and Interest Rates in Europe". *Economic Policy 21*(47), 444–489.

Feng, L., G. Nowak, T. ONeill, and A. Welsh (2014). "CUTOFF: A Spatio-Temporal Imputation Method". *Journal of Hydrology 519*(2014), 3591–3605.

Fjelstul, J. C. and C. J. Carrubba (2018). "The politics of international oversight: Strategic monitoring and legal compliance in the European Union". *American Political Science Review 112*(3), 429–445.

Follesdal, A. and S. Hix (2006). "Why there is a Democratic Deficit in the EU: A Response to Majone and Moravcsik". *Journal of Common Market Studies 44*(3), 533–562.

Fox, J. and S. Weisberg (2011). *An R Companion to Applied Regression* (2 ed.). Thousand Oaks, CA: Sage Publications.

Garbuszus, J. M. and S. Jeworutzki (2016). *"readstata13: Import Stata Data Files"*. R package version 0.8.2.

Gehring, U. W. and C. Weins (2009). *Grundkurs Statistik für Politologen und Soziologen*. Heidelberg: Springer.

Gelman, A. (2006). "Prior Distributions for Variance Parameters in Hierarchical Models". *Bayesian Analysis 1*(3), 515–533.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis* (3 ed.). Boca Raton: CRC press.

Gelman, A. and J. Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

GESIS Leibniz Institute for the Social Sciences (2012a). "Eurobarometer 57.1 - Variable Report". *GESIS-Variable Reports* 2012 | 19, Cologne, GESIS Leibnis Institute for the Social Sciences.

GESIS Leibniz Institute for the Social Sciences (2012b). "Eurobarometer 59.1 - Variable Report". *GESIS-Variable Reports* 2012 | 26, Cologne, GESIS Leibnis Institute for the Social Sciences.

GESIS Leibniz Institute for the Social Sciences (2012c). "Eurobarometer 60.1 - Variable Report". *GESIS-Variable Reports* 2012 | 29, Cologne, GESIS Leibnis Institute for the Social Sciences.

Gill, J. (2008). *Essential Mathematics for Political and Social Research*. Cambridge: Cambridge University Press.

Goldstein, H. (1986). "Multilevel Mixed Linear Model Analysis using Iterative Generalized Least Squares". *Biometrika 73*(1), 43–56.

Google (2018). *Google Scholar*. `https://scholar.google.de/scholar?hl=de&as_sdt=0%2C5&q=binary+time+series+cross+sectional+beck+signorino&btnG=`, accessed on December 18th, 2018.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge: Cambridge University Press.

Greene, W. H. (2012). *Econometric Analysis* (7 ed.). Boston, MA: Pearson.

Grilli, V., D. Masciandaro, and G. Tabellini (1991). "Political and Monetary Institutions and Public Financial Policies in the Industrial Countries". *Economic Policy 6*(13), 341–392.

Grolemund, G. and H. Wickham (2011). "Dates and Times Made Easy with lubridate". *Journal of Statistical Software 40*(3), 1–25.

Hahn, J. (2004). "Does Jeffrey's Prior Alleviate the Incidental Parameter Problem?". *Economics Letters 82*(1), 135–138.

Hallerberg, M. (2002). "Veto Players and the Choice of Monetary Institutions". *International Organization 56*(4), 775–802.

Hallerberg, M. and J. Von Hagen (1999). "electoral institutions, cabinet negotiations, and budget deficits in the european union". In *Fiscal institutions and fiscal performance*, pp. 209–232. Chicago: University of Chicago Press.

Hardy, M. A. (1993). *Regression with Dummy Variables.* Thousand Oaks, CA: Sage Publications.

Hedeker, D. and R. D. Gibbons (2006). *Longitudinal Data Analysis.* Hoboken, NJ: John Wiley & Sons, Ltd.

Heipertz, M. and A. Verdun (2010). *Ruling Europe: The Politics of the Stability and Growth Pact.* Cambridge University Press.

Hix, S. (2007). "Euroscepticism as Anti-Centralization: A Rational Choice Institutionalist Perspective". *European Union Politics 8*(1), 131–150.

Hix, S. and B. Høyland (2011). *The Political System of the European Union.* Basingstoke: Palgrave Macmillan.

Hobolt, S. B. and J. Tilley (2016). "Fleeing the Centre: the Rise of Challenger Parties in the Aftermath of the Euro Crisis". *West European Politics 39*(5), 971–991.

Honaker, J., G. King, and M. Blackwell (2011). "Amelia II: A Program for Missing Data". *Journal of Statistical Software 45*(7), 1–47.

Howarth, D. (2004). "The ECB and the Stability Pact: Policeman and Judge?". *Journal of European Public Policy 11*(5), 832–853.

Hsiao, C., M. H. Pesaran, and A. Pick (2012). "Diagnostic Tests of Cross-Section Independence for Limited Dependent Variable Panel Data Models". *Oxford Bulletin of Economics and Statistics 74*(2), 253–277.

International Monetary Fund (2019a). *General Government Gross Debt. Percent of GDP.* `https://www.imf.org/external/datamapper/GGXWDG_NGDP@WEO/OEMDC/ADVEC/WEOWORLD`, accessed on May 31st, 2019.

International Monetary Fund (2019b). *General Government Net Lending/Borrowing. Percent of GDP.* `https://www.imf.org/external/datamapper/GGXCNL_NGDP@WEO/OEMDC/ADVEC/WEOWORLD`, accessed on May 1st, 2019.

International Monetary Fund (2019c). *Real GDP Growth. Annual Percent Change.* `https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/AZE`, accessed on May 1st, 2019.

Irlenbusch, B., U. Leopold-Wildburger, J. Schütze, and M. Sutter (2003). "Voting in EMU—An Experimental Study of Institutional Innovation and the Role of Communication in the Stability and Growth Pact". *Journal of Common Market Studies 41*(4), 645–664.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* Hoboken, NJ: John Wiley & Sons, Ltd.

Jerez, J. M., I. Molina, P. J. García-Laencina, E. Alba, N. Ribelles, M. Martín, and L. Franco (2010). "Missing Data Imputation Using Statistical and Machine Learning Methods in a Real Breast Cancer Problem". *Artificial Intelligence in Medicine 50*(2), 105–115.

Keele, L. and D. K. Park (2005). "Difficult Choices: an Evaluation of Heterogeneous Choice Models". `https://www.nuffield.ox.ac.uk/Politics/papers/2005/Keele%20Park%20HetChoice%20050315.pdf`, accessed on December 11th, 2018. *(Unpublished Document.).*

King, G. (1989). *Unifying Political Methodology: The Likelihood Method of Statistical Influence.* New York: Cambridge University Press.

King, G., J. Honaker, A. Joseph, and K. Scheve (2001). "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation". *American Political Science Review 95*(1), 49–69.

King, G. and M. E. Roberts (2015). “How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What To Do About It”. *Political Analysis 23*(2), 159–179.

King, G., M. Tomz, and J. Wittenberg (2000). “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”. *American Journal of Political Science 44*(2), 347–361.

King, G. and L. Zeng (2001). “Logistic Regression in Rare Events Data”. *Political Analysis 9*(2), 137–163.

King, S. K. and F. Bouvet (2011). “Interest Rate Setting at the ECB Following the Financial and Sovereign Debt Crises, in Real Time”. *Modern Economy 2*(5), 743.

Köhler, S. and T. König (2015). “Fiscal Governance in the Eurozone: How Effectively Does the Stability and Growth Pact Limit Governmental Debt in the Euro Countries?”. *Political Science Research and Methods 3*(2), 329–351.

Köhler, S., T. König, and X. Lu (2018). “Voter Responsiveness and Responsibility for Supranational Fiscal Governance”. *(Forthcoming)*.

Kohler, U. and F. Kreuter (2008). *Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung* (3 ed.). München, Wien: R. Oldenbourg.

Kong, A., J. S. Liu, and W. H. Wong (1994). “Sequential Imputations and Bayesian Missing Data Problems”. *Journal of the American Statistical Association 89*(425), 278–288.

König, T., D. Finke, and S. Daimer (2005). “Ignoring the Non-Ignorables? Missingness and Missing Positions”. *European Union Politics 6*(3), 269–290.

König, T. and L. Mäder (2014). “The Strategic Nature of Compliance: An Empirical Evaluation of Law Implementation in the Central Monitoring System of the European Union”. *American Journal of Political Science 58*(1), 246–263.

Kontopoulos, Y. and R. Perotti (1999). “government fragmentation and fiscal policy outcomes: Evidence from oecd countries". In *Fiscal institutions and fiscal performance*, pp. 81–102. Chicago: University of Chicago Press.

Kraemer, H. C., G. T. Wilson, C. G. Fairburn, and W. S. Agras (2002). "Mediators and Moderators of Treatment Effects in Randomized Clinical Trials". *Archives of General Psychiatry 59*(10), 877–883.

Kriesi, H. (2007). "The Role of European Integration in National Election Campaigns". *European Union Politics 8*(1), 83–108.

Kuhn, T. and F. Stöckel (2014). "When European Integration Becomes Costly: the Euro Crisis and Public Support for European Economic Governance". *Journal of European Public Policy 21*(4), 624–641.

Lall, R. (2016). "How Multiple Imputation Makes a Difference". *Political Analysis 24*(4), 414–433.

Lane, P. R. (2012). "The European Sovereign Debt Crisis". *Journal of Economic Perspectives 26*(3), 49–68.

Leblang, D. A. (1999). "Domestic Political Institutions and Exchange Rate Commitments in the Developing World". *International Studies Quarterly 43*(4), 599–620.

Lee, D., R. J. Carroll, and S. Sinha (2017). "Frequentist Standard Errors of Bayes Estimators". *Computational Statistics 32*(3), 867–888.

Lee, S.-C. (2016). "A Bayesian Inference for Fixed Effect Panel Probit Model". *Communications for Statistical Applications and Methods 23*(2), 179–187.

Little, R. J. A. (1988a). "Missing-Data Adjustments in Large Surveys". *Journal of Business & Economic Statistics 6*(3), 287–296.

Little, R. J. A. (1988b). "A Test of Missing Completely at Random for Multivariate Data with Missing Values". *Journal of the American Statistical Association 83*(404), 1198–1202.

Long, J. S. (1997). *Advanced Quantitative Techniques in the Social Sciences Series, Vol. 7. Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage Publications.

Lucas Jr, R. E. and N. L. Stokey (1983). "Optimal Fiscal and Monetary Policy in an Economy Without Capital". *Journal of Monetary Economics 12*(1), 55–93.

Maas, C. J. M. and J. J. Hox (2004). “Robustness Issues in Multilevel Regression Analysis”. *Statistica Neerlandica 58*(2), 127–137.

Mahto, A. (2018). *“splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values"*. R package version 1.4.4.

Majone, G. (1998). “Europe’s ’Democratic Deficit’: The Question of Standards”. *European Law Journal 4*(1), 5–28.

Majone, G. (2001). “Nonmajoritarian Institutions and the Limits of Democratic Governance: a Political Transaction-Cost Approach”. *Journal of Institutional and Theoretical Economics JITE 157*(1), 57–78.

Mansfield, E. D. and E. Reinhardt (2003). “Multilateral Determinants of Regionalism: The Effects of GATT/WTO on the Formation of Preferential Trading Arrangements”. *International Organization 57*(4), 829–862.

Marien, S. and M. Hooghe (2011). “Does Political Trust Matter? An Empirical Investigation Into the Relation between Political Trust and Support for Law Compliance”. *European Journal of Political Research 50*(2), 267–291.

Martin, A. D. (2008). *Bayesian Analysis*. Oxford: Oxford University Press.

McCulloch, C. E. (2003). “Generalized Linear Mixed Models”. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, Volume 7, pp. i–84. Beachwood, OH: Institute of Mathematical Statistics.

McLaren, L. (2007). “Explaining Mass-Level Euroscepticism: Identity, Interests, and Institutional Distrust”. *Acta Politica 42*(2-3), 233–251.

Moravcsik, A. (2002). “Reassessing Legitimacy in the European Union”. *Journal of Common Market Studies 40*(4), 603–624.

Nordhaus, W. D. (1975). “The Political Business Cycle". *The Review of Economic Studies 42*(2), 169–190.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley & Sons.

OECD (2017). *List of OECD Member Countries - Ratification of the Convention on the OECD.* http://www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm, accessed on December 11th, 2018.

OSCE (2019). *Elections in Estonia.* https://www.osce.org/odihr/elections/estonia, accessed on July 14th, 2019.

Pang, X. (2010). "Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-Sectional Data: A Bayesian Multilevel Model with AR (p) Errors". *Political Analysis 18*(4), 470–498.

Pearl, J. (2009). "Causal Inference in Statistics: An Overview". *Statistics Surveys 3*(2009), 96–146.

Pennock, J. R. (1952). "Responsiveness, Responsibility, and Majority Rule". *American Political Science Review 46*(3), 790–807.

Persson, T. and L. E. Svensson (1989). "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences". *The Quarterly Journal of Economics 104*(2), 325–345.

Pierce, C. A., R. A. Block, and H. Aguinis (2004). "Cautionary Note on Reporting Eta-Squared Values from Multifactor ANOVA Designs". *Educational and Psychological Measurement 64*(6), 916–924.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC". *R News 6*(1), 7–11.

Protocol (No 12) on the Excessive Deficit Procedure (2008). *Consolidated Version of the Treaty on European Union - PROTOCOLS - Protocol (No 12) on the Excessive Deficit Procedure.* http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008M/PRO/12&from=EN, accessed on December 11th, 2018.

R Core Team (2015). *"foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...".* R package version 0.8-66.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005). "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects". *Journal of Econometrics 128*(2), 301–323.

Raffalovich, L. E. and R. Chung (2014). "Models for Pooled Time-Series Cross-Section Data". *International Journal of Conflict and Violence 8*(2), 210.

Reinhart, C. M., V. R. Reinhart, and K. S. Rogoff (2012). "Public Debt Overhangs: Advanced-Economy Episodes since 1800". *Journal of Economic Perspectives 26*(3), 69–86.

Reinhart, C. M. and K. S. Rogoff (2010). "Growth in a Time of Debt". *American Economic Review 100*(2), 573–78.

Rodden, J. (2002). "The dilemma of fiscal federalism: Grants and fiscal performance around the world. *American Journal of Political Science 46*(3), 670–687.

Roubini, N. and J. D. Sachs (1989). "Political and Economic Determinants of Budget Deficits in the Industrial Democracies". *European Economic Review 33*(5), 903–933.

Rubin, D. B. (2005). "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions". *Journal of the American Statistical Association 100*(469), 322–331.

Rudolph, T. J. and J. Evans (2005). "Political Trust, Ideology, and Public Support for Government Spending". *American Journal of Political Science 49*(3), 660–671.

Schafer, J. L. (1999). "Multiple Imputation: a Primer". *Statistical Methods in Medical Research 8*(1), 3–15.

Schafer, J. L. and J. W. Graham (2002). "Missing Data: Our View of the State of the Art". *Psychological Methods 7*(2), 147–177.

Shah, A. D., J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway (2014). "Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: a CALIBER Study". *American Journal of Epidemiology 179*(6), 764–774.

Shor, B., J. Bafumi, L. Keele, and D. Park (2007). "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data". *Political Analysis 15*(2), 165–181.

Singer, J. D. and J. B. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Stéclebout-Orseau, E. and M. Hallerberg (2007). "Strategic Signals and Multilateral Fiscal Surveillance". `http://aei.pitt.edu/7993/`, accessed on December 11th, 2018. *(Unpublished Document.)*.

Stegmüller, D. (2013). "How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches". *American Journal of Political Science 57*(3), 748–761.

Stekhoven, D. J. and P. Bühlmann (2012). "MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data". *Bioinformatics 28*(1), 112–118.

Su, Y.-S. and M. Yajima (2015). *"R2jags: Using R to Run 'JAGS'"*. R package version 0.5-7.

Tabellini, G. (1991). "The Politics of Intergenerational Redistribution". *Journal of Political Economy 99*(2).

Teddlie, C. and A. Tashakkori (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences.* Thousand Oaks, CA: Sage Publications.

Teorell, J., S. Dahlberg, S. Holmberg, B. Rothstein, N. A. Pachon, and R. Svensson. (2019). *The Quality of Government Standard Dataset 2019. Codebook.* University of Gothenburg: The Quality of Government Institute. `https://www.qogdata.pol.gu.se/data/qog_std_jan19.pdf`, accessed on May 1st, 2019.

Thatcher, M. and A. S. Sweet (2002). "Theory and Practice of Delegation to Non-Majoritarian Institutions". *West European Politics 25*(1), 1–22.

The World Bank Group (2019). *GDP (constant 2010 US $).* `https://data.worldbank.org/indicator/ny.gdp.mktp.kd`, accessed on May 1st, 2019.

Tierney, N., D. Cook, M. McBain, and C. Fay. *"naniar: Data Structures, Summaries, and Visualisations for Missing Data"*. R package version 0.4.0.9000.

Van Buuren, S. *Flexible Imputation of Missing Data. Second Edition., edition = 2, year=2018, publisher=Boca Raton, FL: Chapman and Hall/CRC.*

van Buuren, S. v. and K. Groothuis-Oudshoorn (2011). "mice: Multivariate Imputation by Chained Equations in R". *Journal of Statistical Software 45*(3), 1–68.

van der Veer, R. A. and M. Haverland (2018). "Bread and Butter or Bread and Circuses? Politicisation and the European Commission in the European Semester". *European Union Politics 0*(0), 1–22.

Vanberg, G. (2001). "Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review". *American Journal of Political Science 45*(2), 346–361.

Vanberg, G. (2005). *The Politics of Constitutional Review in Germany*. Cambridge: Cambridge University Press.

VanderWeele, T. J. and J. M. Robins (2007). "Four Types of Effect Modification: A Classification Based on Directed Acyclic Graphs". *Epidemiology 18*(5), 561–568.

Velasco, A. (2000). Debts and deficits with fragmented fiscal policymaking. *Journal of Public Economics 76*(1), 105–125.

von Hagen, J. (2002). Fiscal rules, fiscal institutions, and fiscal performance. *The Economic and Social Review 33*(3), 263–284.

von Hagen, J. (2006). "Political Economy of Fiscal Institutions". In B. R. Weingast and D. A. Wittman (Eds.), *The Oxford Handbook of Political Economy*, pp. 464–478. Oxford: Oxford University Press.

von Hagen, J. and I. J. Harden (1995). "Budget Processes and Commitment to Fiscal Discipline". *European Economic Review 39*(1995), 771–779.

von Hagen, J. and G. B. Wolff (2006). "What Do Deficits Tell Us About Debt? Empirical Evidence on Creative Accounting with Fiscal Rules in the EU". *Journal of Banking & Finance 30*(12), 3259–3279.

von Hippel, P. T. (2009). "How to Impute Interactions, Squares, and Other Transformed Variables". *Sociological Methodology 39*(1), 265–291.

Wallace, T. D. and A. Hussain (1969). "The Use of Error Components Models in Combining Cross Section with Time Series Data". *Econometrica 37*(1), 55–72.

Warnes, G. R., B. Bolker, G. Gorjanc, G. Grothendieck, A. Korosec, T. Lumley, D. MacQueen, A. Magnusson, J. Rogers, et al. (2017). *"gdata: Various R Programming Tools for Data Manipulation"*. R package version 2.18.0.

Weingast, B. R., K. A. Shepsle, and C. Johnsen (1981). "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics". *Journal of Political Economy 89*(4), 642–664.

Western, B. and S. Jackman (1994). "Bayesian Inference for Comparative Research". *American Political Science Review 88*(2), 412–423.

Wibbels, E. (2000). "Federalism and the Politics of Macroeconomic Policy and Performance". *American Journal of Political Science 44*(4), 687–702.

Wickham, H. (2007). "Reshaping Data with the reshape Package". *Journal of Statistical Software 21*(12), 1–20.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer.

Wickham, H. (2017). *"scales: Scale Functions for Visualization".* R package version 0.5.0.

Wickham, H. and J. Bryan (2017). *"readxl: Read Excel Files".* R package version 1.0.0.

Wickham, H., R. François, L. Henry, and K. Müller (2017). *"dplyr: A Grammar of Data Manipulation".* R package version 0.7.4.

Wickham, H. and L. Henry (2018). *"tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions".* R package version 0.8.0.

Wijsman, S. and C. Crombez (2017). "The Political Economy of Enforcing Fiscal Rules". *FEB Research Report* MSI_1701, Leuven: KU Leuven - Faculty of Economics and Business.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach.* (4 ed.). Mason, OH: South-Western Cengage Learning.

Yatchew, A. and Z. Griliches (1985). "Specification Error in Probit Models". *The Review of Economics and Statistics 67*(1), 134–139.