The Politics of External Approval: Explaining the IMF's Evaluation of Austerity Programs

This is the accepted version of

Abstract

During the European debt crisis, numerous states launched austerity programs. The International Monetary Fund (IMF) evaluates and forecasts the likelihood of member states' success in implementing these programs. Although IMF evaluations influence country risk perceptions on capital markets, little is known about their reasoning. This paper uses fuzzy-set Qualitative Comparative Analysis to explore on what grounds the IMF evaluated the success prospects of austerity programs during the European debt crisis. Results reveal that IMF evaluations are heavily influenced by the program's implementation credibility. They require a tractable policy problem, a country's institutional capacity to structure implementation, and favor expenditure reduction over revenue measures. By acting as a strict guide on the road to fiscal adjustment, the IMF indirectly influences member states' scope of policymaking through its surveillance activities. Extensive austerity programs that need to be implemented swiftly are evaluated negatively if the country is not involved in an IMF program.

Keywords:

Introduction

This paper analyzes on what grounds the IMF evaluated the success prospects of austerity programs during the European debt crisis. The level of debt in developed European countries has risen sharply in recent years. To avoid spiraling debt and refinance existing debts on the capital markets under acceptable conditions, numerous states have adopted austerity programs. Austerity programs aim to balance the fiscal budget by reducing expenditure and increasing revenue, hence constituting one way to achieve fiscal consolidation (Blyth 2013; IMF 1995). The International Monetary Fund (IMF), often described as one of the most powerful IOs in history (Bird 2007; Nelson 2014; Stone 2002), "accompanies" countries on the road to fiscal consolidation. Throughout this process, it influences domestic fiscal policy in important ways (Ban 2015; Barnett & Finnemore 2004; Fang & Stone 2012; Woods 2006). Besides providing policy advice to member countries, the IMF also evaluates the countries' likely performance in implementing austerity policies, and communicates the results (Dreher, Marchesi & Vreeland 2008; IMF 2015a, b, c; Lombardi & Woods 2008). The IMF's evaluations of countries' austerity programs – a key element in the IMF's bilateral surveillance activities – are influential signals to capital markets on how risky certain sovereigns are in terms of investment (Ban & Gallagher 2015; Dreher et al. 2008; IMF 2015c). If the IMF does not consider a country's road to fiscal adjustment credible, this results in an increased risk perception on capital markets (Fratzscher & Reynaud 2011).

Despite the huge relevance of IMF austerity program evaluations, little is known about their reasoning. The main thrust of research on IMF-decision-making has focused on the Fund's lending activities, but paid little attention to surveillance activities (Barnett & Finnemore 2004; Dreher & Gassebner 2012; Pop-Eleches 2008; Steinwand & Stone 2008; Stone 2004). However, as the IMF's new *raison d'être* is surveillance (Dreher et al. 2008), the question of how the IMF exerts power on its member states – not only by direct means such as formal

enforcement mechanisms, but also indirectly through the narrowing of policy advice and options – has taken on a new dimension (Broome & Seabrooke 2012).

As arguably the first study that examines the IMF's recent stance towards austerity in the context of its surveillance activities in the Eurozone, this paper asks: how did the IMF assess austerity programs during the European debt crisis? We adopt a public policy perspective and assume that Sabatier and Mazmanian's (1980) seminal framework of the implementation process helps us understand IMF evaluations (e.g., Exadaktylos & Zahariadis 2014). Employing fuzzy-set Qualitative Comparative Analysis (fsQCA) (Ragin 2008), we assess how different constellations of the interplay of problem tractability, the ability of the austerity program to structure implementation, and non-statutory variables affected the IMF's evaluation of 20 austerity programs implemented in 14 countries of the Eurozone or the European Union (EU) during the recent debt crisis.

This paper contributes to the debate on IMF evaluations in three ways. First, it analyzes developed European countries. Earlier work has focused on low- and middle-income transition and developing countries and emphasized the IMF's responsiveness to geopolitical and transnational business interest (Aldenhoff 2007; Dreher et. al. 2008; Fratzscher & Reynaud 2011; Pop-Eleches 2009; Thacker 1990). This perspective has largely neglected the "domestic variables that influence whether or not a government adopts an IO's policy preferences" (Broome & Seabrooke 2012: 2). Indeed, findings on the IMF's loan negotiations and lending decisions (Broome 2010; Chwieroth 2013; Stone 2002; Woods 2006) suggest that particularly in the institutionally and politically more stable context of the Eurozone, domestic variables should affect how the IMF evaluates austerity programs (Heller 2002). Second, contrary to these previous studies, we focus on IMF evaluations within the context of formal assessments of austerity programs which provide information signals about sovereign risk to capital markets. Finally, our analysis focuses on recently issued austerity programs. Since 2008, the IMF began

to reconsider its understanding of a "sound fiscal policy" and its stance towards austerity and stimulus measures. It increasingly recommended a more gradual implementation of austerity measures for countries with sufficient confidence on financial markets (Blanchard & Leigh 2013) and a more balanced mix of expenditure reduction and revenue increases (Ban 2015). Given this altered stance towards austerity, the assumption that the IMF acts as a fierce and undifferentiated agent of austerity made in much work on the European debt crisis may prove to be too narrow. Clearly, a more differentiated stance towards austerity by the IMF asks for an in-depth examination of the factors that influence that stance.

In fact, our results show that the IMF is not an undifferentiated agent of austerity, but considers domestic conditions. IMF evaluations prove to be heavily influenced by implementation credibility. First, for rather uncompetitive countries with lower institutional capacity and ambitious austerity programs, it is almost impossible to take the road to fiscal adjustment without being involved in an IMF rescue program. Second, the IMF negatively evaluates economically stronger countries whose adjustment policies do not conform to the IMF's preference of expenditure reduction over revenue measures. These findings suggest that the IMF acts as a strict guide on the road to fiscal adjustment, indirectly influencing member states' scope of policymaking through its surveillance activities.

Next, we discuss the IMF's evaluations of austerity programs and outline our explanatory model. We then proceed to the research design and methods employed. After presenting the results, we go on to discuss the main findings and their implications.

The Politics of External Approval

The IMF has recently begun to gradually shift its primary focus from lending to surveillance activities. Accordingly, scholars seek to understand how surveillance works and what implications it has for member countries. Particular focus has been placed on the forecasts of

4

fiscal and macroeconomic variables of member countries which are published on a regular basis.

Although IMF forecasts are known to be based on a strong scientific culture (Ban & Gallagher 2015), the IMF publishes only vague information on these procedures and schematically describes the preceding consultation processes (Asdorian 2015; IMF 2015b, c): "the initial projections are based on an econometric model. Subsequently, however, there is much leeway for […] discretionary adjustments" (Dreher et al. 2008: 146). IMF evaluations contain a substantive qualitative element resulting from the considerable subjectivity of the respective country teams (IMF 2015b, c). While earlier research scrutinized the accuracy of forecasts (Pons 2000; Timmermann 2007), few scholars have recently analyzed their determinants (Aldenhoff 2007; Dreher et. al. 2008; Fratzscher & Reynaud 2011). These studies have found a low explanatory power of political strategies and macroeconomic variables, while quite unambiguously suggesting that the IMF both legitimizes its lending activities with overly optimistic forecasts, and favors countries that are political allies of the United States of America (US) (Aldenhoff 2007; Dreher et al. 2008). In emerging market economies, the countries' political power has also proven relevant (Fratzscher & Reynaud 2011).

We argue that in the context of the contemporary Eurozone, domestic variables that influence a country's implementation credibility matter for IMF austerity program evaluations. While the geopolitical and strategic interests of the IMF's principals set the broader boundaries for IMF-decision-making (Stone 2002; Thacker 1990), within these boundaries, the IMF has acquired considerable maneuvering space (Chwieroth 2013; Pop-Eleches 2009; Woods 2006). Whereas IMF evaluations may be commercially more sensitive towards systemic economic entities like the Eurozone than to low- and middle-income transition and developing economies, this factor alone cannot explain the considerable variation in IMF austerity program evaluations *within* the Eurozone. Geopolitical arguments also cannot be the main difference-maker in the European

context: the IMF has negatively rated the austerity programs of powerful US-friendly countries like Germany and the United Kingdom. Finally, the IMF has expressed its intention to strengthen the traction of surveillance by ensuring its quality, candor, and evenhandedness (IMF 2011: 19; 2014), which should also enhance the relevance of domestic variables.

When a country announces an austerity program, this can be interpreted as a signal of its willingness for fiscal adjustment. IMF publications clearly indicate that the IMF does not take these signals at face value, but subjects them to a "reality-check" to verify them for their implementation credibility (Heller 2002). Our aim is to evaluate the IMF's response to the announcement of an austerity program, specifically, whether the IMF deems the program capable of meeting its objective. The IMF regularly publishes forecasts for estimated budget balances in its *Fiscal Monitor (FM)* (Ban & Gallagher 2015; IMF 2015a, b). The outcome we seek to explain is the IMF's evaluation of an austerity program's likely success in balancing the budget as intended, hereinafter referred to as "positive evaluation" (POS). To capture this, we compare the fiscal balance targeted by the austerity program, on the one hand, with the fiscal balance (general government structural balance) projection after the announcement of an austerity program, on the other.[1]

-- insert Figure 1 here –

---

[1] To ensure that announced austerity programs were actually incorporated in the projections, we chose a time span of approximately 4-6 weeks between announcement and data publication by the IMF. In case of doubt, we opted for the later Fiscal Monitor update.

As Figure 1 illustrates, the IMF's forecasts of changes in the countries' fiscal balance after implementation vary considerably between different austerity programs. If the IMF forecasts that the program is likely to miss the target, then the evaluation is more negative than positive. For instance, by the end of 2010, Ireland announced its intention to meet the EU deficit criteria of -3% of GDP until 2014. However, considering the planned austerity measures, the IMF projected the deficit to be at -4.8%. Hence, the IMF expected Ireland to miss its target by 1.8 percentage points. If the IMF projects that the program will not miss its intended target (score of 0 or more), then the program is more positively than negatively evaluated. The Slovak Republic, for instance, announced an austerity program in May 2012, with the aim of bringing its budget deficit down to 3% by 2013. In response, the IMF expected the program to exceed its goal by +0.1 percentage points and projected the deficit to be 2.9% after implementation. The observed evaluations range from +1,4 to -3.7. It is these differences that we now scrutinize.

On what grounds can the IMF reasonably assess?

When announcing austerity programs, countries must, as a minimum, define the total amount of savings, the expected duration, and cost saving efforts in order to obtain funds and reasonable credit terms. However, most austerity programs under scrutiny outlined these parameters too vaguely to ensure their successful implementation. The IMF was therefore well-advised to treat these indications as a mere declaration of intent, whose implementation credibility required verification. Sabatier and Mazmanian's (1980) framework captures the way austerity measures as centrally designed top-level policy decisions, often imposed under external pressure, are implemented particularly well (e.g., Exadaktylos & Zahariadis 2014). We adapt that framework to the specific research context by integrating explanatory factors mentioned more or less explicitly by the IMF. Sabatier and Mazmanian (1980) understand the implementation of public policies to be the result of the interplay between three main sets of factor. First, the tractability

of the problem at hand is captured here as program ambitiousness and the presence of an IMF rescue program. Second, the ability of the statute to structure implementation is influenced by the number of veto players and the effectiveness of the national administration. Third, we consider the presence of a strong Centre-Right government and a country's economic competitiveness as non-statutory variables affecting implementation.

Sabatier and Mazmanian's (1980: 554) model should be understood as a "minimum list of crucial conditions", rather than as individual factors which work in isolation. We want to discover case-specific *configurations* that explain how the IMF evaluates austerity programs. Surprising results then provide opportunities for further explorations to refine theory (Rihoux & Ragin 2009). While adopting an explorative focus, we seek to identify theoretically plausible counterfactual arguments when dealing with limited diversity (that not all logically possible combinations of relevant causal conditions exist in the real world). To this end, we formulate directional expectations on the effects of single variables. These expectations serve as counterfactual arguments and are not testable hypotheses (Schneider & Wagemann 2012: 295ff).

Problem tractability

First, the greater the amount of behavioral change required, the more problematic a program's successful implementation becomes (Sabatier & Mazmanian, 1980: 544). The *ambitiousness of the adopted program* (AMB) sets the size of the envisaged deficit reduction in relation to the time countries had at their disposal to achieve the deficit reduction. Programs are ambitious if the country envisaged a large budget reduction within a brief period of time. This might indicate low implementation credibility to the IMF: ambitious programs could be mere "paper tigers" that promise an unrealistically severe and swift fiscal change. Conversely, the IMF might also value high ambitiousness as the appropriate austerity measure, indicating that a country takes

its austerity plans seriously. Since program ambitiousness could indicate both high and low success prospects, no directional expectation is formulated.

Second, two reasons make us assume that the IMF rates austerity programs more positively in countries where it *is engaged with a rescue program* (PRG) (IMF 2015b). First, there is strong evidence that the IMF engages in "defensive forecasting": a negative evaluation would weaken the credibility of the IMF program. Positive evaluations can also increase the probability that those loans are repaid (Aldenhoff 2007; Dreher et al. 2008). Second, the presence of an IMF program signals a country's willingness and ability to undertake substantive reform (Fratzscher & Reynaud 2011: 407). Countries with IMF arrangements in place have already undergone a 'screening' process that allows the IMF to better estimate the country's ability to fulfill its policy commitments (Chwieroth 2013). Less uncertainty could positively affect IMF evaluations. Moreover, loan disbursements during an IMF program are conditional on demonstrable policy actions, which provides the IMF with opportunities to directly influence, monitor and enforce policy design and compliance (Armingeon & Guthmann, 2014; Broome, 2015; IMF 2015d).

Ability of the statute to structure implementation

Third, implementation success is determined by the number of veto/clearance points involved in the attainment of statutory objectives (Sabatier & Mazmanian, 1980: 546). Austerity programs can entail severe and abrupt cuts affecting a wide range of players and giving rise to opposition. Central governments must make sure that "budgetary discipline is imposed on all fiscal entities" (Heller 2002: 18). A decentralized political system features a larger number of sub-national veto players whose opposition to painful austerity measures could potentially undermine the "flexible and responsive execution of the budget" (Huber, Ragin & Stephens 1993; IMF 1995: 34; Tsebelis 1995). Thus, a *high number of veto players in a country* (DC)

should negatively affect the IMF's evaluation of implementation credibility.

Fourth, sufficient financial and personnel administrative resources are "necessary to hire the staff and to conduct the technical analyses involved in the development of regulations, the administration of permit programs, and the monitoring of compliance" (Sabatier & Mazmanian 1980: 545). Established administrative capacities, e.g., a strong treasury, improve the efficient, thorough and successful implementation of austerity programs (Heller 2002: 17; IMF 1995). The *existence of an effective administration* (EFF) should thus positively affect the IMF's evaluation of the austerity program.

Non-statutory variables

Fifth, Sabatier and Mazmanian (1980: 574) highlight that "[a]ny new program requires implementors who are not merely neutral but sufficiently persistent to develop new regulations and standard operating procedures, and to enforce them in the face of resistance from target groups". IMF evaluations entail a continuous "policy dialogue" between fund staff and national authorities (Lombardi & Woods 2008). Much of the existing literature has focused on the commitment of individual 'sympathetic interlocutors', i.e. "national policymakers who are sympathetic to advice from the IMF" (Broome & Seabrooke 2015: 6; Chwieroth 2015). We measure the political commitment of the whole government through partisan affiliation for three reasons. First, in developed countries, the IMF pays more attention to broader political support than to the presence of individual interlocutors (Woods 2006). Second, in situations like the European debt crisis marked by a very contentious economic problem and alternative ideology-based interpretations and solutions, partisanship influences the IMF's decision-making (Pop-Eleches 2009). Third, in a European context key actors are not only socialized by the universities they attend, but also by their political parties. Due to their political commitment to a less comprehensive welfare state policy, conservative Centre-Right governments typically

support and implement austerity measures to a stronger degree than Centre-Left and Christian-Democratic governments (Hibbs 1977). The former usually face less blame when cutting social policy (Giger & Nelson 2010). Accordingly, the *presence of a strong Centre-Right government* (CR) should positively affect the IMF's evaluation of implementation credibility.[2]

Sixth, as Bird (2007: 683) points out, "in order to understand the IMF's operations, economics has to be combined with politics". A relevant socio-economic condition is the competitiveness of the economy. Economic competitiveness describes the basic conditions for growth, innovation, and efficiency that determine how productively a country uses its available resources to provide future economic prosperity. Economically competitive states are more likely able to escape the debt spiral by generating economic growth and reducing both their debt burden and future pressure for savings. This, in turn, creates budgetary scope and increases the likelihood that budgetary targets are met (IMF 2015a). Thus, the countries' *high economic competitiveness* (COM) should positively influence the IMF's evaluation of success likelihood.

Table 1 summarizes the six conditions and the directional expectations used for counterfactual arguments. Adopting QCA notation, the presence of a factor is indicated with uppercase letters, and its absence with lowercase letters.

-- insert Table 1 here --

---

[2] We account for the "policy dialogue" influencing the IMF's assessment of political commitment during our case discussions below. During regular Article IV meetings, the IMF tries to grasp the political commitment of country officials for announced fiscal policies (Lombardi & Woods 2008).

Data and methods

We assess 20 austerity programs introduced in 14 countries within the Eurozone or the EU since the crisis began between January 2009 and May 2012.[3] This design holds important economic contextual factors constant: the countries share an internal market and have been undergoing a process of fiscal-political harmonization since the crisis began. Since they have the same currency, or their national currency is tied with the Euro, they cannot avoid consolidation measures by devaluing their currency without restriction.[4] This enhances the importance of austerity programs.

We employ Fuzzy Set Qualitative Comparative Analysis (fsQCA) (Ragin 2008; Rihoux & Ragin 2009; Schneider & Wagemann 2012) to identify necessary and/or sufficient conditions for a positive or negative evaluation by the IMF (software: QCA R package). The underlying assumption of causal complexity has three elements. *Equifinality* means that various scenarios can induce the same IMF evaluation. *Conjunctural causation* indicates that case-specific factors affect IMF evaluation in combination rather than in isolation (Sabatier and Mazmanian 1980). Lastly, *asymmetrical causation* means that different causal factors may matter for a positive IMF evaluation than for a negative IMF evaluation. Case knowledge enables us to identify potential measurement errors (Schneider & Wagemann 2012: 89, 295-305, 307-312).

fsQCA conceives of variables as sets in which cases have membership or not. The attribution of cases to sets is called calibration. Qualitative anchors determine the stage at which the

[3] In the absence of any official document listing austerity programs passed in the EU, cases were identified using Internet-based research. National and international news services were consulted to verify that the announced austerity program had actually been adopted.
[4] This includes the UK or Denmark, who cannot avoid internal devaluation at the expense of currency devaluation.

condition is deemed fully present (fuzzy value ≥ 0.95), fully absent (fuzzy value ≤ 0.05) and an indifference point at 0.5. The latter establishes the difference in kind: for example, fuzzy membership values in POS above 0.5 means that IMF evaluation was rather or fully positive (POS), while values below 0.5 indicate that the evaluation was rather or fully negative (pos).

fsQCA uses the logical operators OR (+) and AND (*). The latter depicts combinations of conditions, referred to hereinafter as configurations or paths. A "truth table" shows all possible combinations of conditions. If all or enough cases' fuzzy set membership in a truth table row is smaller than or equal to its membership in the outcome, then the row is identified as a sufficient path for the outcome. The logical minimization process identifies the shortest expression for those factors that imply (→) the outcome – the solution term.

fsQCA results are evaluated using two main parameters of fit that range from 0-1. The appropriate levels for these parameters are research-specific but are better the closer they are to 1 (Schneider & Wagemann 2012: 128). *Consistency* indicates the extent to which the results are in line with the statements of necessity or sufficiency, which is weakened by "deviant cases consistency in kind" with qualitatively different membership in the explanation and the outcome (Schneider & Rohlfing 2013). The proportional reduction in inconsistency (PRI) indicates the degree to which a configuration is not simultaneously sufficient for both outcomes. Consistency should not be below 0.75 for sufficient conditions, and 0.9 for necessary conditions (Ragin 2008: 46). The presence of "gaps" and deviant cases consistency in kind helped us determine raw consistency thresholds for single truth table rows. *Coverage* then states how well the available empirical information is explained by the condition(s). For sufficient conditions, raw coverage indicates how much a single path covers, while unique coverage indicates how much it uniquely covers. For necessary conditions, coverage expresses their relevance in terms of not being much larger than the outcome, and the Relevance of Necessity (RoN), in terms of the condition being close to a constant (Schneider & Wagemann 2012: 128, 139, 235-239).

Applying the Enhanced Standard Analysis (ESA) procedure, we make theoretically informed directional expectations about empirically unobserved configurations (Table 1) and ensure that the coding of the outcome in the truth table does not contradict prior findings of necessity or sufficiency (Schneider & Wagemann, 2012: 198-211). The datasets, descriptive statistics, the truth tables, directional expectations, conservative and parsimonious solution terms, simplifying assumptions and r codes for replication are all indicated in the online appendix.

Operationalization and calibration

We now turn to the measurement and calibration of the condition and outcome sets, using the current values at the time of the announcement of the austerity program (Table 2). The calibration decisions are outlined in detail together with extensive robustness tests in online appendix B (Skaaning 2011).

Highly ambitious program (AMB). Austerity programs are ambitious if they need to achieve a large deficit reduction within too little time to be realistically implemented. To measure ambitiousness, we divide the planned deficit reduction in per cent of GDP by the duration of the austerity program in months. For instance, while Portugal's first austerity program sought to bring the deficit down from -7.3% to -4.6% of GDP almost within a year, the Czech program aimed at reducing the deficit from -3.5% to 0% within four years. The resulting scale has a sample range from 0.6 (very high ambitiousness) to 0.01 (very low ambitiousness).

Existence of an IMF program (PRG). Austerity programs in countries with an IMF program in place at the time of the announcement are coded with 1, all others, with 0 (dichotomous set).

Strongly decentralized political system (DC). In decentralized political systems, sub-national jurisdictions have significant decision-making power concerning policy design and implementation and can make significant changes (e.g., in the timetable and broad parameters). By contrast, in centralized systems, sub-national units cannot significantly influence the design

and implementation of austerity measures, which results only in minor to austerity policies adjustments (e.g., local delays). The aggregated decentralization index developed by the Assembly of European Regions views decentralization as the sum of decision-making power and competencies of sub-national jurisdictions in various areas (scale from 0 to 100) (AER 2009).

Effective state administration (EFF). An effective state administration possesses the means and expertise to implement austerity measures thoroughly, on time and successfully. This is not the case if it lacks important resources and expertise necessary to translate austerity requirements into concrete policies and enforce their implementation. The Governance Indicator developed by Kaufmann, Kraay and Mastruzzi (2011) combines the quality of the public service, the political independence of the administration, the quality of policy formulation and implementation and other administration-related aspects into an index for government effectiveness that ranges from -2.5 to +2.5. Higher values correspond to greater government effectiveness.

Strong Centre-Right government (CR). If the share of government-posts held by Centre-Right candidates is sufficiently large, Centre-Right candidates' can exert a dominant influence on policymaking and implementation. We use the percentage share of (ministerial) posts held by Centre-Right parties in the government from the Comparative Political Data Set III 1990-2010 of Armingeon et al. (2012) for operationalization.

High economic competitiveness (COM). A high degree of competitiveness facilitates sound future growth and economic prosperity, as increased tax revenue creates budgetary leeway and alleviates the need for austerity in subsequent years. Conversely, low economic competitiveness heralds poor growth prospects and significantly limits the future budgetary scope. The Global Competitiveness Index of the World Economic Forum aggregates data on basic conditions for growth, efficiency-boosting factors and innovation-boosting factors on a numerical scale which

can range from 2.7 to 5.8.

-- insert Table 2 here --

Results

Results reveal two necessary conditions for a positive evaluation by the IMF (for full results see Table A4 and Figure A1 online appendix). First, the IMF requires a country to possess the institutional capacity to structure the implementation of austerity measures in order to consider the latter feasible: whenever the IMF gave an austerity program a positive evaluation, either the political system tended to be centralized, or the state administration was effective (dc + EFF).

A second prerequisite for the IMF to evaluate an austerity program positively is a tractable problem: either the austerity program has to be non-ambitious, or an ongoing IMF program is in place (amb + PRG). The existence of an IMF program might actually influence the program's perceived ambitiousness: it reduces information asymmetries and enables the IMF to enforce maximum compliance instead of having to rely on persuasion (Fang & Stone 2012; IMF 2015b).

Table 3 presents the three paths that imply that the IMF grades austerity programs well. The single cases that are explained by this solution, the consistency and coverage indicators for the single paths and the overall solution are listed below. Cases can display several paths. We discuss typical cases for each path below.

-- insert Table 3 here --

It is striking that a comparatively centralized political system is a necessary part of the story for

a positive evaluation. Paths 1 and 2 describe the "struggling, but manageable students": austerity programs passed in countries with ineffective state administrations (eff), but which are comparatively centralized (dc) and with a direct involvement of the IMF (PRG). In the first path, this combined with low program ambitiousness (amb), and in path 2, with a strong Centre-Right government (CR). In line with the "defensive forecasting" assertion, these cases are Greece and Portugal, where the IMF provided financial support and was directly involved in the implementation of agreed austerity measures. The IMF seems to be confident that its engagement in a country can bolster a weak state administration under favorable conditions marked by fewer veto points, reasonable consolidation targets, or a Centre-Right government that supports the austerity program. Whether this is due to the IMF's closer grip on implementation, the improved availability of information, or a combination of these, could be addressed in future research by interviewing IMF officials about their motives.

Another group of positively evaluated austerity programs entailed "model pupils" with comparatively unambitious programs (amb) adopted in economically competitive (COM), centralized countries (dc) with a strong Centre-Right government (CR) and an effective administration (EFF). These ideal conditions for the successful implementation of austerity measures were present in Denmark and Finland. For instance, Finland devised a program which was aimed merely at "keeping the balance". Since there was never any need to reestablish Finland's standing on capital markets, the existence of an IMF program was irrelevant for the IMF's evaluation.

Figure 2 illustrates how the cases score on this solution. Despite the low coverage, we are able to explain six out of nine cases of positive evaluation. France is a deviant case, as we will discuss below.

-- insert Figure 2 here --

Another four paths imply that the IMF evaluated austerity programs negatively (Table 4). Three factors have a particular salience when the IMF rates austerity programs negatively: first, the programs should not be overly ambitious; second, low economic competitiveness is often a hurdle; and third, the IMF is obviously skeptical when it is not engaged in a country.

-- insert Table 4 here --

Both paths 1 and 2 suggest that in the absence of direct IMF involvement, the IMF considers ambitious programs to be "likely deceivers". Examples for path 1 are the two austerity programs passed in quick succession in Italy under the Berlusconi government and under the Monti government, which entailed severe cuts to be implemented swiftly (AMB), but facing significant veto power in a decentralized country (DC) without IMF involvement (prg). Low problem tractability, Italy's low ability to structure implementation, and the IMF's low control over the latter lent little credibility to these austerity programs. As an instance for path 2, in May 2010, Spain announced that it wanted to reduce its deficit by 3.3 percentage points in no more than 19 months (AMB). The highly ambitious program did not convince the IMF: the IMF had no direct control or information (prg), political support by the Leftist government under Zapatero was weak (cr), and Spain was comparatively uncompetitive economically (com).

The third pathway to a negative evaluation of implementation credibility entails "hopeless cases" like the Czech Republic's 2012 austerity program with a particularly unfavorable configuration of conditions, namely: strong decentralization (DC), low effectiveness of state administration (eff), low economic competitiveness (com), and no IMF involvement (prg). Understandably, the IMF deems the successful implementation of such programs unlikely,

irrespective of how ambitious they are.

The fourth path covers the Irish austerity program as a special, "overly ambitious" case: it envisaged a deficit reduction of no less than 28.9 percentage points in 48 months. The IMF projected that the program would miss its target by "only" -1.8 points. The presence of a rather effective state administration (EFF) combined with low economic competitiveness (com), but the program's exceptionally high ambitiousness was probably decisive for the IMF. Indeed, the evaluation could be seen as rather positive. The solution has a high explanatory power, covering 8 out of 11 negative evaluations (Figure 3).

-- insert Figure 3 here --

Our proposed framework had a limited capacity to explain the evaluation of austerity programs in Western European countries like Austria, Belgium, France and Germany, but also Slovakia. Figure 1 has already revealed that, given the wide range of IMF evaluations, regional provenience cannot itself be the missing explanation. Based on the IMF country reports, we now discuss such "outlier" cases (Rihoux & Ragin 2009; Schneider & Rohlfing 2013) (see Table A9 online appendix).

Austria's two austerity programs represent the "least explained cases" for negative and positive evaluation. As they display an identical country context and the same configuration of explanatory factors, we can identify the decisive additional factor through a controlled comparison. Austria announced its first austerity program, which aimed at reducing the deficit by 2 % within 47 months, in January 2010. The IMF criticized Austria for over-emphasizing revenue measures in its plan for fiscal consolidation, and projected that it would miss the target by -0.9%, due to negative effects on growth and sustainability. The second Austrian austerity program was launched in February 2012 to lower the deficit by 0.6 % within 58 months. The

IMF projected that this program would exceed its target by 1.4 %: Austria had proven responsive to the IMF's calls to shift the focus of fiscal policy to the reduction of expenditures, now striking an appropriate balance between fiscal discipline and the cyclical needs of the economy.

This comparison suggests that the IMF evaluates countries more positively if they emphasize expenditure reduction for achieving fiscal balance, especially during the initial stages of adjustment (Broome 2015; IMF 1995: 26). Indeed, the IMF also reacted to the other unexplained negatively rated programs with calls for more expenditure containment (France 2012, Germany 2010). Conversely, the other positively rated deviant cases (Belgium 2012, Slovak Republic 2012) were explicitly lauded by the IMF for their focus on expenditure containment.

The French austerity program, announced in August 2011, was rated negatively, despite the fact that France is a "model pupil". The IMF predicted a slight miss of -0.3 per cent (targeted reduction: 1.4 per cent in 16 months). The Danish 2010 program and the Finnish 2012 program belonged to the same group of "model pupils" and received a positive evaluation. Here, the crux of the matter was tax policy. The IMF staff pointed out that the very high French tax burden would keep increasing relative to France's peers. Conversely, the IMF welcomed the tax freeze implemented in Denmark and lauded the Finish taxation measures as steps in the right direction.

These comparisons suggest that, particularly in the context of Western European countries, the IMF considers not only implementation credibility, but also the austerity programs' content – specifically, their relative emphasis on revenue measures or expenditure reduction. As austerity programs often outline this content in broad terms only, the IMF appears also to take past experience into account (Heller 2002).

Conclusions

This paper investigated the grounds on which the IMF assessed the prospects for success of austerity programs during the European debt crisis. The methodological approach we adopted differs from the conventional toolbox of IMF studies, as it focuses on interrelated and substitutable, rather than isolated net effects on IMF surveillance. This in turn paves the way for an analysis of the causal mechanisms through which IOs influence their member states. Simultaneously, our less-than-intermediate-N fuzzy set analysis entails relatively high amounts of limited diversity and precludes highly generalizable results, or statements about the size of the causal effects. Extensive robustness tests and case knowledge helped us control for measurement error (Skaaning 2011).

Despite these limitations, using Sabatier and Mazmanian's (1980) seminal framework, our study is the first to show that the austerity programs' domestic implementation credibility heavily influences IMF surveillance (Broome & Seabrooke 2012). First, the IMF requires a minimum of *institutional capacity to structure implementation* to give an austerity program a positive evaluation. Second, the *problem has to have a minimum level of tractability*, either because the program is not very ambitious, or because an IMF rescue program is in place. Indeed, third, the *IMF's engagement in a country* appears to be crucial. The IMF took a more skeptical stance when it was not directly involved, particularly if austerity programs were overly ambitious and/or the unfavorable implementation context made failure likely. Finally, the IMF favored *expenditure reduction over revenue measures*. These findings cohere with the relevance of domestic conditions for IMF-decision-making within the context on conditionality (Broome 2010; Chwieroth 2013; Stone 2002; Woods 2006).

This paper has departed from much research on IMF-decision-making by examining the specific context of IMF surveillance in developed European countries. One of our main motivations was that insights on IMF-decision-making in one context may not be readily

transferable to another. We focused on comparable cases which share an internal market, where the absence of the option of devaluating the currency enhances the importance of austerity programs. In many developing countries, where economic policy is more centralized and less transparent than in the European context, the IMF usually gets a closer grip on key actors and agencies responsible for implementation (Woods 2006: 82). Accordingly, individual 'sympathetic interlocutors' should be given more attention by the IMF when dealing with developing countries (Broome & Seabrooke 2015; Chwieroth 2015; Woods 2006). Moreover, considering complementarities between budget policy and institutional configurations in different varieties of capitalism may be a promising avenue for future research (Amable & Azizi 2014; Hall & Soskice 2001).

Within the scope of our sample, our results suggest that the IMF is a strict guide on the road to fiscal adjustment. Particularly in cases where austerity requirements are substantial and must be implemented swiftly, the IMF tends not to give positive evaluations if the country is not involved in an IMF program. While these patterns are compatible with the "defensive forecasting" argument (Aldenhoff 2007; Dreher et al. 2008; Fratscher & Reynaud 2011), our overall results suggest that this is as much about implementation credibility and, ultimately, about exerting influence as it is about legitimizing lending activities. First, when the IMF's policy advice is backed up with financial incentives for domestic compliance through loan programs, it is harder for national authorities to reject the IMF's reform recommendations (Broome 2015: 149). Second, the IMF possesses more information on and control over whether a government is serious about enacting "IMF friendly" policy reforms (Broome & Seabrooke 2012: 2).

To date, the literature has examined IMF influence predominantly within the context of lending activities (Bird 2007; Barnett & Finnemore 2004; Dreher & Gassebner 2012; Steinwand & Stone 2008; Stone 2002, 2004). However, our study illustrates that evaluations are a relatively

obvious way for the IMF to effectively exert indirect influence on member states via its surveillance activities. First, for rather uncompetitive countries with lower institutional capacity, it is almost impossible to take the road to fiscal adjustment without the IMF. By tendency, countries that need to restore their standing on international capital markets can do so only under the IMF's direct supervision. If countries reject this help, they pay for their independence from the IMF with higher country-risk perceptions on capital markets. A higher risk perception on capital markets heralds rising refinancing costs, which, in turn, implies diminished scope for independent policymaking (Ban & Gallagher 2015). Hence, weak countries in need of fiscal consolidation pay the price in terms of sovereignty not only after they receive funding from the IMF, but already through the IMF's evaluation of their fiscal position. Second, regardless of any existing IMF program, the IMF also uses its evaluations to impose its preferred policy options upon economically less vulnerable countries. This calls for further research examining the realm of indirect influence that exists between the IMF and its member countries.

References

Aldenhoff, F.O. (2007). Are economic forecasts of the International Monetary Fund politically biased? A public choice analysis. *Review of International Organizations, 2(3),* 239-260.

Amable, B. & Azizi, K. (2014). Counter-cyclical budget policy across varieties of capitalism. *Structural Change and Economic Dynamics, 30,* 1-9.

Armingeon, K. & Guthmann, K. (2014). Democracy in crisis? The declining support for national democracy in European countries, 2007–2011. *European Journal of Political Research 53,* 423-442.

Armingeon, K., Careja, R., Weisstanner, D., Engler, S., Potolidis, P. & Gerber, M. (2012). *Comparative Political Data Set III 1990-2010.* Institute of Political Science, University of Bern 2012.

Asdorian, G. (2015). *International Monetary Fund World Economic Outlook Data Forum.* Retrieved from http://forums.imf.org/showthread.php?162248-Information-on-Staff-Estimates

Assembly of European Regions (AER) (2009). *From subsidiarity to success: The Impact of Decentralisation on Economic Growth.* Retrieved from http://www.aer.eu/publications/aer-study.html

Ban, C. (2015). Austerity versus Stimulus? Understanding Fiscal Policy Change at the International Monetary Fund Since the Great Recession. *Governance, 28(2),* 167-183.

Ban, C. & Gallagher, K. (2015). Recalibrating Policy Orthodoxy: The IMF Since the Great Recession. *Governance,* 28(2), 131-146.

Barnett, M. & Finnemore, M. (2004). *Rules for the World – International Organizations in Global Politics.* Ithaca: Cornell University Press.

Bird, G. (2007). The IMF: A bird's eye view of its role and operations. *Journal of Economic Surveys, 21(4),* 683-745.

Blanchard, O. & Leigh, D. (2013). Growth Forecast Errors and Fiscal Multipliers. *IMF Working Paper, 13(1),* 1-42.

Blyth, M. (2013). *Austerity: The History of a Dangerous Idea.* Oxford: Oxford University Press.

Broome, A. (2010). *The Currency of Power: The IMF and Monetary Reform in Central Asia.* New York: Palgrave Macmillan.

Broome, A. (2015). Back to Basics: The Great Recession and the Narrowing of IMF Policy Advice. *Governance, 28(2),* 147-165.

Broome, A. & Seabrooke, L. (2012). Seeing Like an International Organisation. *New Political Economy, 17(1),* 1-16.

Broome, A. & Seabrooke, L. (2015). Shaping Policy Curves: Cognitive Authority in Transnational Capacity Building. *Public Administration,* early view.

Chwieroth, J.M. (2013). "The silent revolution": How the staff exercise informal governance over IMF lending. *Review of International Organizations, 8(2),* 265-290.

Chwieroth, J.M. (2015). Professional ties that bind: how normative orientations shape IMF conditionality. *Review of International Political Economy, 22(4),* 757-787.

Dreher, A., Marchesi, S. & Vreeland, J.R. (2008). The political economy of IMF forecasts.

Public Choice, 137, 145-171.

Dreher, A. & Gassebner, M. (2012). Do IMF and World Bank Programs Induce Government Crises? An Empirical Analysis. *International Organization, 66(2),* 329-358.

Exadaktylos, T. & Zahariadis, N. (2014). Quid pro Quo: Political Trust and Policy Implementation in Greece during the Age of Austerity. *Politics & Policy, 42(1),* 160-183.

Fang, S. & Stone, R.W. (2012). International organizations as policy advisors. *International Organization, 66(4),* 537-569.

Fratzscher, M. & Reynaud, J. (2011). IMF surveillance and financial markets - A political economy analysis. *European Journal of Political Economy, 27,* 405-422.

Giger, N. & Nelson, M. (2010). The electoral consequences of welfare state retrenchment: Blame avoidance or credit claiming in the era of permanent austerity? *European Journal of Political Research, 50(1),* 1-23.

Hall, P.A. & Soskice, D. (2001). Varieties of Capitalism. The Institutional Foundations of Comparative Advantage. Oxford. Oxford University Press.

Heller, P.S. (2002). Considering the IMF's Perspective on a 'Sound Fiscal Policy'. *IMF Policy Discussion Paper, 02/08,* 1-27.

Hibbs, D.A. Jr., (1977). Political Parties and Macroeconomic Policy. *The American Political Science Review, 71(4),* 1467-1487.

Huber, E., Ragin, C.C. & Stephens, J.D. (1993). Social Democracy, Christian Democracy, Constitutional Structure, and the Welfare State. *American Journal of Sociology, 99(3),* 711-749.

International Monetary Fund (IMF) (1995). *Guidelines for Fiscal Adjustment. Pamphlet Series No. 49.* Retrieved from https://www.imf.org/external/pubs/cat/longres.aspx?sk=620.0

International Monetary Fund (IMF) (2011). *2011 Triennial Surveillance Review – Overview Paper.* Retrieved from http://www.imf.org/external/np/pp/eng/2011/082911.pdf.

International Monetary Fund (IMF) (2014). *2014 Triennial Surveillance Review – Overview Paper.* Retrieved from http://www.imf.org/external/np/pp/eng/2014/073014.pdf.

International Monetary Fund (IMF) (2015a). *Technical assistance and Training. International Monetary Fund Factsheet.* Retrieved from http://www.imf.org/external/np/exr/facts/tech.htm

International Monetary Fund (IMF) (2015b). *The IMF and Europe. International Monetary Fund Factsheet.* Retrieved from http://www.imf.org/external/np/exr/facts/europe.htm

International Monetary Fund (IMF) (2015c). *IMF Surveillance. International Monetary Fund Factsheet.* Retrieved from http://www.imf.org/external/np/exr/facts/surv.htm

International Monetary Fund (IMF) (2015d). *IMF Conditionality. International Monetary Fund Factsheet.* Retrieved from https://www.imf.org/external/np/exr/facts/conditio.htm

International Monetary Fund (IMF). *Country reports.* Retrieved from http://www.imf.org/external/country/

Kaufmann, D., Kraay, A. & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law, 3,* 220-246.

Lombardi, D. & Woods, N. (2008). The politics of influence: An analysis of IMF surveillance.

Review of International Political Economy, 15(5), 711-739.

Nelson, S.C. (2014). The International Monetary Fund's evolving role in global economic governance. In M. Moschella & C. Weaver (eds), *Handbook of Global Economic Governance: Players, Power and Paradigms.* New York: Routledge.

Organisation for Economic Co-operation and Development (OECD) (2013). Country statistical profile: Czech Republic. *Country statistical profiles: Key tables from OECD.* Retrieved from http://www.oecd-ilibrary.org/economics/country-statistical-profile-czech-republic-2013_csp-cze-table-2013-1-en;jsessionid=18y5bf37xvut7.x-oecd-live-02

Pons, J. (2000). The Accuracy of IMF and OECD Forecasts for G7 Countries. *Journal of Forecasting, 19(1),* 53-63.

Pop-Eleches, G. (2008). Crisis in the Eye of the Beholder: Economic Crisis and Partisan Politics in Latin American and East European International Monetary Fund Programs. *Comparative Political Studies, 41(9),* 1179-1211.

Pop-Eleches, G. (2009). From Economic Crisis to Reform: IMF Programs in Latin America and Eastern Europe. Princeton, NJ: Princeton University Press.

Ragin, C.C. (2008). *Redesigning Social Inquiry - Fuzzy Sets and Beyond.* Chicago: University of Chicago Press.

Rihoux, B. & Ragin, C.C. (2009). *Configurational comparative methods. Qualitative Comparative Analysis (QCA) and related techniques.* Thousand Oaks and London: Sage Publications.

Sabatier, P. & Mazmanian, D. (1980). The Implementation of Public Policy: A Framework of Analysis. *Policy Studies Journal, 8(4),* 538-560.

Schneider, C.Q. & Rohlfing, I. (2013). Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research. *Sociological Methods & Research, 42(4),* 559-597.

Schneider, C.Q. & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis.* New York: Cambridge University Press.

Skaaning, S. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods & Research, 40(2),* 391-408.

Steinwand, M.C. & Stone, R.W. (2008). The International Monetary Fund: A review of the recent evidence. *Review of International Organizations, 3,* 123-149.

Stone, R.W. (2002). *Lending Credibility: The International Monetary Fund and the Post-Communist Transition.* Princeton, NJ: Princeton University Press.

Stone, R.W. (2004). The political economy of IMF lending in Africa. *American Political Science Review, 98(4),* 577-592.

Thacker, S.C. (1990). The High Politics of IMF Lending. *World Politics, 52(1),* 38-75.

Timmermann, A. (2007). An evaluation of the world economic outlook forecasts. *IMF Staff Papers, 54(1),* 1-33.

Tsebelis, G. (1995). Decision Making in Political Systems: Veto Players in Presidentialism, Parlamentarism, Multicameralism and Multipartism. *British Journal of Political Science, 25(3),* 289-325.

Woods, N. (2006). *The globalizers: the IMF, the World Bank, and their borrowers.* Cornell University Press.

World Economic Forum (WEF). *Competitiveness Report(s) 2008-2009, 2009-2010, 2010-*

2011, 2011-2012. Retrieved from http://www.weforum.org/reports/global-competitiveness-report-2011-2012

Tables and Figures

Table 1: Conditions and directional expectations

Condition	*Ceteris paribus, condition produces positive evaluation (POS) when…*	*Ceteris paribus, condition produces negative evaluation (pos) when…*
Problem tractability		
Highly ambitious program		
AMB	no expectation	no expectation
Existence of an IMF program		
PRG	present	absent
Ability of the statute to structure implementation		
Highly decentralized political system		
DC	Absent	present
Effective state administration		
EFF	Present	absent
Non-statutory variables		
Strong Centre-Right government		
CR	Present	absent
High economic competitiveness		
COM	Present	absent

Note: directional expectations denote counterfactual arguments rather than empirically testable hypotheses (Schneider & Wagemann 2012: 168-177)

Table 2: Measurement and calibration

	Set	Measurement	*Fully out* 0.05	*Neither in nor out* 0.5	*Fully in* 0.95
			Calibration (set membership)		
Outcome	Positive evaluation of cost-saving measures by IMF (POS)	Difference between budget deficit reduction targeted by austerity program and budget deficit reduction forecast by IMF (Fiscal Monitor)	-2.4	-0.15	1.4
	Highly ambitious program (AMB)	Required reduction in fiscal balance in per cent, divided by duration of program in months	0.01	0.1205	0.20
	Existence of an IMF program (PRG)	Dichotomous variable	0	--	1
	Highly decentralized political system (DC)	AER Decentralization Index (2009)	33.5	47	55
	Effective state administration (EFF)	Government Effectiveness partial index of the Governance Indicator of Kaufmann et al. (2011)	0.7	1.175	2.065
	Strong Centre-Right government (CR)	Percentage share of Centre-Right parties in government (Comparative Political Data Set III 1990-2010 by Armingeon et al. (2012))	5	52	95
	High economic competitiveness (COM)	WEF Global Competitiveness Index	4.28	4.86	5.285

Table 3: Sufficient conditions for positive evaluation

Intermediate solution	amb***PRG***dc***eff** + **PRG***dc***eff***CR + **amb***dc***EFF***CR***COM → POS		
Single case coverage			
	GRC3; GRC2,PRT2	GRC2,PRT2; GRC1	DNK,FIN, *FRA*
Consistency	0.908	0.796	0.848
Raw coverage	0.228	0.269	0.247
Unique coverage	0.050	0.092	0.243
Solution consistency 0.831; Solution coverage 0.563			

Bold: parsimonious solution (direct causal interpretability).   *Italics*: deviant case consistency in kind.

Table 4: Sufficient conditions for negative evaluation

Intermediate solution	**AMB*****prg***DC + **AMB*****prg***cr*com + prg***DC*****eff***com + **AMB*****EFF***com → pos			
Single case coverage	ITA2,ITA3; GBR;ESP	PRT1;ESP	ESP;CZE; ITA1	IRL
Consistency	0.969	1.000	0.946	1.000
Raw coverage	0.421	0.212	0.404	0.212
Unique coverage	0.099	0.084	0.084	0.084
Solution consistency 0.931; Solution coverage 0.643				

Bold: parsimonious solution (direct causal interpretability).   *Italics*: deviant case consistency in kind.

Figure 1: IMF evaluations of austerity programs, January 2009 - May 2012

IMF evaluation
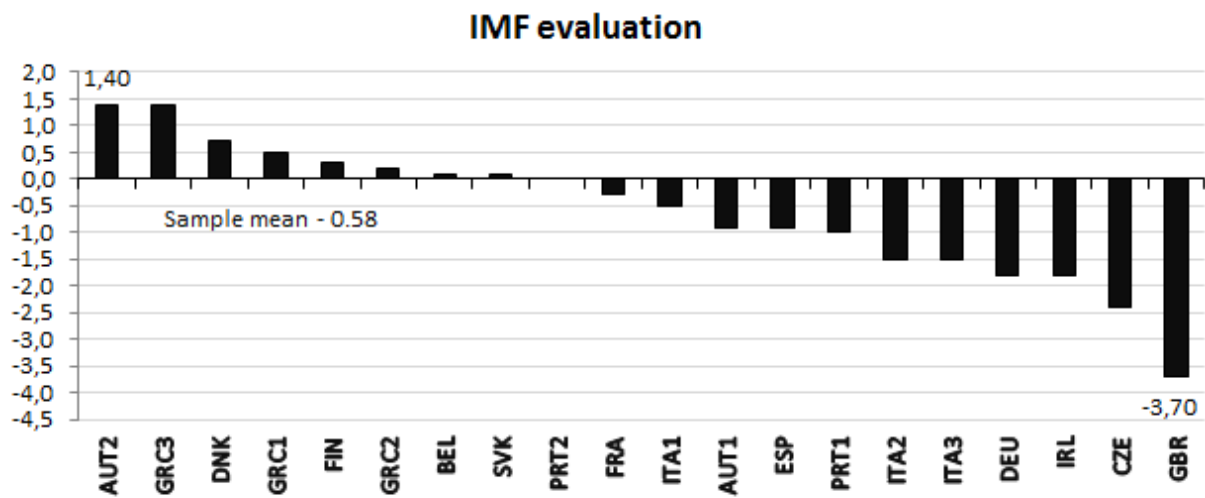
Key: AUT = Austria, BEL = Belgium, CZE = Czech Republic, DEU = Germany, DNK = Denmark, ESP = Spain, FIN = Finland, FRA = France, GBR = United Kingdom, GRC = Greece, IRL = Ireland, ITA = Italy, PRT = Portugal, SVK = Slovak Republic.

Figure 2: Intermediate solution for positive evaluation
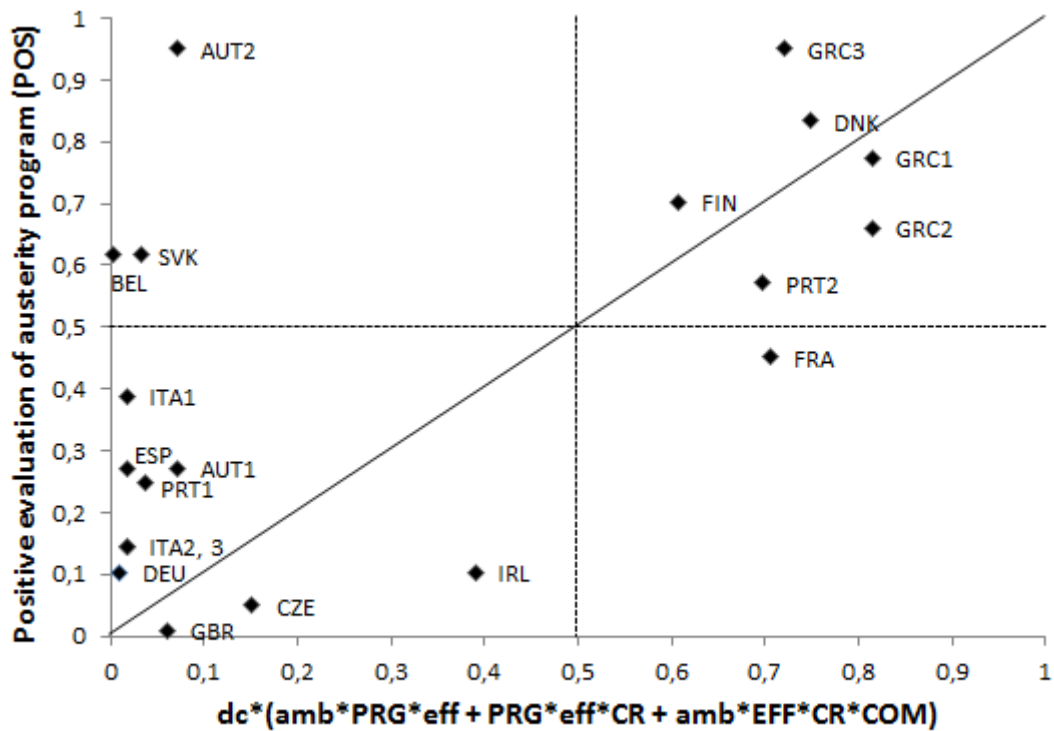
Cases situated above the diagonal are consistent. In the upper left quadrant are deviant cases for coverage, in the lower right quadrant are deviant cases consistency in kind. The lower left quadrant is irrelevant (Schneider & Rohlfing 2013).

Figure 3: Intermediate solution for negative evaluation
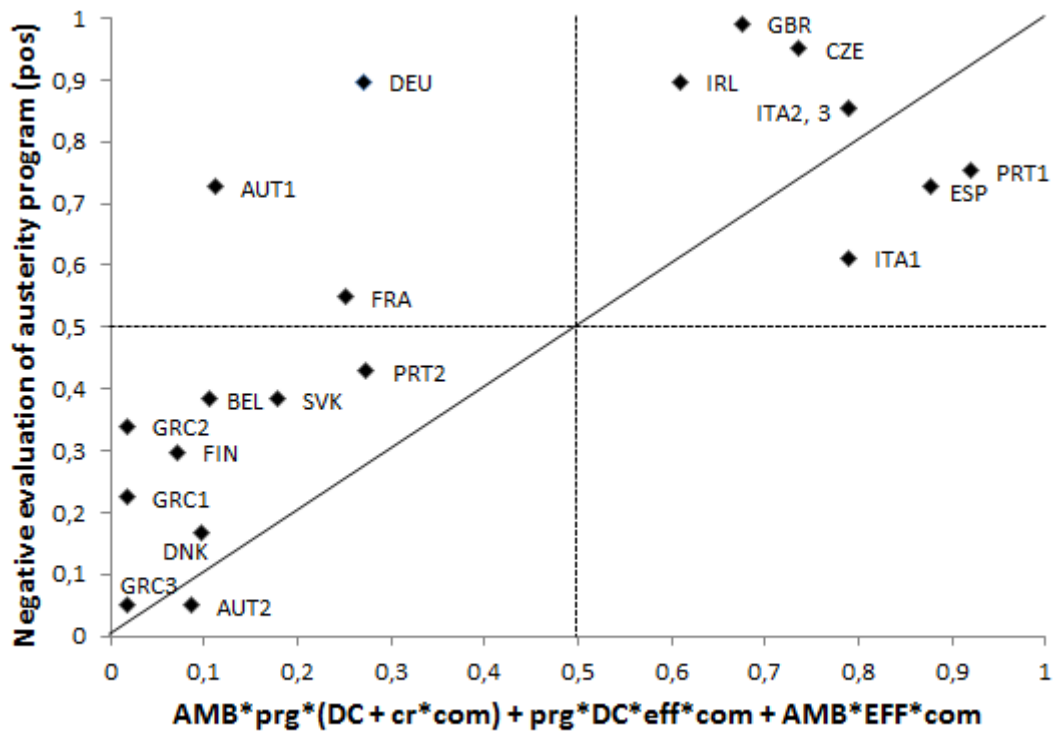
Cases situated above the diagonal are consistent. In the upper left quadrant are deviant cases for coverage, in the lower right quadrant are deviant cases consistency in kind. The lower left quadrant is irrelevant (Schneider & Rohlfing 2013).

Online Appendix A: Supplementary tables and figures

Table A1: Raw data

Case ID	Country	Date of Announcement	Duration until	POS	Required reduction (%)	Duration (months)	Ambitiousness	DC	EFF	CR	COM	PRG
AUT1	Austria	01/2010	2013	-0.9	2	47	0.042553191	54	1.89	50	5.13	0
AUT2	Austria	02/2012	2016	1.4	0.6	58	0.010344828	54	1.89	50	5.14	0
BEL	Belgium	03/2012	2012	0.1	0.4	10	0.04	63	1.59	54.54	5.2	0
CZE	Czech Republic	04/2012	2016	-2.4	3.5	56	0.0625	50	1.01	45.45	4.52	0
DEU	Germany	06/2010	2014	-1.8	4.5	54	0.083333333	60	1.55	100	5.37	0
DNK	Denmark	05/2010	2013	0.7	1.6	43	0.037209302	42	2.29	100	5.46	0
ESP	Spain	05/2010	2011	-0.9	3.3	19	0.173684211	58	0.98	0	4.59	0
FIN	Finland	03/2012	2016	0.3	1.4	57	0.024561404	45	2.24	100	5.47	0
FRA	France	08/2011	2012	-0.3	1.4	16	0.0875	42	1.44	88.41	5.13	0
GBR	United Kingdom	05/2010	2014	-3.7	10.5	54	0.194444444	49	1.56	100	5.19	0
GRC1	Greece	05/2010	2014	0.5	10.6	55	0.192727273	31	0.52	73.73	4.04	1
GRC2	Greece	05/2011	2015	0.2	4.4	54	0.081481481	31	0.52	73.73	3.99	1
GRC3	Greece	02/2012	2015	1.4	3.9	46	0.084782609	31	0.52	36.84	3.92	1
IRL	Ireland	12/2010	2014	-1.8	28.9	48	0.602083333	41	1.31	86.28	4.74	1
ITA1	Italy	09/2010	2012	-0.5	2.1	26	0.080769231	50.6	0.52	100	4.37	0
ITA2	Italy	08/2011	2013	-1.5	4.3	28	0.153571429	50.6	0.52	100	4.37	0
ITA3	Italy	12/2011	2013	-1.5	4	24	0.166666667	50.6	0.52	100	4.43	0
PRT1	Portugal	11/2010	2011	-1	2.7	13	0.207692308	42	1.04	0	4.38	0
PRT2	Portugal	05/2011	2014	0	2.6	31	0.083870968	42	1.04	66.67	4.4	1
SVK	Slovak Republic	05/2012	2013	0.1	1.2	19	0.063157895	36	0.88	0	4.19	0

Survey period: July/August 2012.

Table A2: Descriptive statistics of raw variables

Variable	Minimum	Maximum	Mean	Median	Standard deviation	Skew
Pos	-3.70	1.40	-0.58	-0.40	1.29	-0.49
Amb	0.01	0.60	0.12	0.08	0.13	2.52
Prg	0.00	1.00	0.25	0.00	0.44	1.07
Dc	31.00	63.00	46.14	47.00	9.49	-0.08
Eff	0.52	2.29	1.19	1.04	0.59	0.35
Cr	0.00	100.00	66.28	73.73	35.62	-0.70
Com	3.92	5.47	4.70	4.55	0.51	0.09

N = 20.

Table A3: Fuzzy data

Case	POS	AMB	DC	EFF	CR	COM	PRG
AUT1	0.27	0.11	0.93	0.91	0.47	0.87	0.00
AUT2	0.95	0.05	0.93	0.91	0.47	0.87	0.00
BEL	0.62	0.10	1.00	0.80	0.54	0.91	0.00
CZE	0.05	0.18	0.75	0.26	0.40	0.15	0.00
DEU	0.10	0.27	0.99	0.78	0.96	0.97	0.00
DNK	0.83	0.10	0.25	0.98	0.96	0.98	0.00
ESP	0.27	0.88	0.98	0.23	0.04	0.20	0.00
FIN	0.70	0.07	0.39	0.97	0.96	0.99	0.00
FRA	0.45	0.29	0.25	0.71	0.92	0.87	0.00
GBR	0.01	0.94	0.68	0.78	0.96	0.91	0.00
GRC1	0.77	0.94	0.03	0.02	0.82	0.02	1.00
GRC2	0.66	0.26	0.03	0.02	0.82	0.01	1.00
GRC3	0.95	0.28	0.03	0.02	0.28	0.01	1.00
IRL	0.10	1.00	0.21	0.61	0.91	0.35	1.00
ITA1	0.39	0.26	0.79	0.02	0.96	0.08	0.00
ITA2	0.15	0.77	0.79	0.02	0.96	0.08	0.00
ITA3	0.15	0.85	0.79	0.02	0.96	0.10	0.00
PRT1	0.25	0.96	0.25	0.30	0.04	0.08	0.00
PRT2	0.57	0.27	0.25	0.30	0.73	0.09	1.00
SVK	0.62	0.18	0.08	0.14	0.04	0.03	0.00

Table A4: Analysis of necessity

Condition	Positive evaluation (POS)			Negative evaluation (pos)		
	Consistency	*Coverage*	*RoN*	*Consistency*	*Coverage*	*RoN*
AMB	0.432	0.437	0.695	0.661	0.840	0.889
DC	0.519	0.442	0.623	0.719	0.769	0.800
EFF	0.573	0.578	0.751	0.510	0.647	0.783
CR	0.746	0.500	0.505	0.768	0.647	0.592
COM	0.531	0.550	0.748	0.465	0.604	0.771
PRG	0.345	0.612	0.885	0.174	0.388	0.831
amb	0.842	0.664	0.699	0.558	0.552	0.635
Dc	0.729	0.674	0.769	0.478	0.556	0.710
eff	0.650	0.514	0.617	0.667	0.662	0.699
Cr	0.473	0.618	0.836	0.407	0.668	0.854
com	0.617	0.479	0.590	0.654	0.637	0.673
prg	0.655	0.387	0.352	0.826	0.613	0.463
dc + EFF	**0.971**	**0.610**	**0.517**	-	-	-
amb + PRG	**0.959**	**0.608**	**0.523**	-	-	-
cr + amb[1]	**0.906**	**0.601**	**0.555**	-	-	-
AMB + DC[1]	-	-	-	**0.909**	**0.735**	**0.63**

Bold: Condition passes consistency threshold of 0.9.

[1]No necessary condition: at least one deviant case consistency in kind.

A consistent necessary condition is deemed trivial if coverage is below 0.6 and the RoN value is below 0.5. We tested for all possible supersets of POS and pos. Only complex necessary conditions meeting the consistency and triviality criteria are listed in the table.

Table A5: Truth table for outcome "POS"

AMB	DC	EFF	CR	COM	PRG	POS	Consistency	PRI	Cases
0	0	0	0	0	1	1	1.000	1.000	GRC3
0	0	0	1	0	1	1	0.885	0.771	GRC2,PRT2
0	0	1	1	1	0	1	0.841	0.691	DNK,FIN,FRA
1	0	0	1	0	1	1	0.838	0.704	GRC1
0	1	1	0	1	0	0	0.819	0.589	AUT1,AUT2
0	1	1	1	1	0	0	0.720	0.438	BEL,DEU
0	0	0	0	0	0	0	0.718	0.371	SVK
1	1	0	0	0	0	0	0.633	0.001	ESP
0	1	0	1	0	0	0	0.607	0.041	ITA1
1	0	0	0	0	0	0	0.585	0.001	PRT1
0	1	0	0	0	0	0	0.582	0.058	CZE
1	1	1	1	1	0	0	0.519	0.000	GBR
1	0	1	1	0	1	0	0.458	0.000	IRL
1	1	0	1	0	0	0	0.434	0.000	ITA2,ITA3

Raw consistency threshold: 0.837 (next highest 0.819; AUT1 is a deviant case consistency in kind).

Complex solution: amb*dc*eff*com*PRG + dc* CR*eff*com*PRG + amb*dc* CR*EFF*COM*prg → POS (solution consistency 0.831, solution coverage 0.563).

Parsimonious solution (without exclusion of untenable assumptions): eff*PRG + amb*dc*CR → POS (solution consistency 0.782, solution coverage 0.640).

Untenable assumptions: dc*EFF + AMB*prg → POS (contradicts statement of necessity).

Enhanced parsimonious solution (under exclusion of untenable assumptions): amb*dc* CR + dc*eff*PRG → POS (solution consistency 0.786, solution coverage 0.640).

Limited diversity: 50 out of 64 configurations are logical remainders (78.2%). Directional expectations see Table 1.

Table A6: Simplifying assumptions for analysis of POS

AMB	DC	EFF	CR	COM	PRG	Easy counterfactual (used for intermediate solution?
0	0	0	0	1	1	X
0	1	0	0	0	1	-
0	1	0	0	1	1	-
1	0	0	0	0	1	-
1	0	0	0	1	1	-
1	1	0	0	0	1	-
1	1	0	0	1	1	-
0	0	0	1	0	0	-
0	0	0	1	1	0	-
0	0	0	1	1	1	X
0	0	1	1	0	0	-
0	0	1	1	0	1	-
0	0	1	1	1	1	X
0	1	0	1	0	1	-
0	1	0	1	1	1	-
1	0	0	1	1	1	X
1	1	0	1	0	1	-
1	1	0	1	1	1	-

Table A7: Truth table for outcome "pos"

AMB	DC	EFF	CR	COM	PRG	pos	Consistency	PRI	Cases
1	0	1	1	0	1	1	1.000	1.000	IRL
1	1	0	1	0	0	1	1.000	1.000	ITA2,ITA3
1	1	1	1	1	0	1	1.000	1.000	GBR
1	0	0	0	0	0	1	1.000	0.999	PRT1
1	1	0	0	0	0	1	0.976	0.934	ESP
0	1	0	0	0	0	1	0.974	0.942	CZE
0	1	0	1	0	0	1	0.923	0.811	ITA1
0	1	1	1	1	0	0	0.782	0.562	BEL,DEU
0	1	1	0	1	0	0	0.740	0.411	AUT1,AUT2
0	0	0	0	0	0	0	0.700	0.329	SVK
1	0	0	1	0	1	0	0.594	0.259	GRC1
0	0	1	1	1	0	0	0.583	0.190	DNK,FIN,FRA
0	0	0	1	0	1	0	0.496	0.000	GRC2,PRT2
0	0	0	0	0	1	0	0.458	0.000	GRC3

Raw consistency threshold: 0.0.922 (next highest 0.782; BEL is a deviant case consistency in kind).

Complex solution: DC*eff*com*prg + AMB*eff*cr*com*prg + AMB*dc*EFF*CR*com*PRG + AMB*DC*EFF* CR*COM*prg → pos (solution consistency 0.965, solution coverage 0.638).

Parsimonious solution (without exclusion of untenable assumptions):

The present data display tied logically redundant prime implicants and hence, a certain degree of ambiguity. Both models are reported below (Schneider & Wagemann, 2012, pp. 108ff). The two solutions are identical, except for the role of com and eff, respectively in the last path. We opt for M2 because of its higher consistency and coverage. The intermediate solution is identical for both models.

M1: AMB*EFF + AMB*prg + DC*com → pos (solution consistency 0.925, solution coverage 0.682).

M2: AMB*EFF + AMB*prg + DC*eff → pos (solution consistency 0.930, solution coverage 0.687).

Untenable assumptions: amb*dc* CR + dc*eff*PRG → pos (contradicts statement of sufficiency for POS, enhanced parsimonious solution).

Enhanced parsimonious solution (under exclusion of untenable assumptions): amb*dc* CR + dc*eff*PRG → POS (solution consistency 0.786, solution coverage 0.640).

M1: AMB*EFF + AMB*prg + DC*com → pos (solution consistency 0.925, solution coverage 0.682).

M2: AMB*EFF + AMB*prg + DC*eff → pos (solution consistency 0.930, solution coverage 0.687).

Limited diversity: 50 out of 64 configurations are logical remainders (78.2%). Directional expectations see Table 1.

Table A8: Simplifying assumptions for analysis of pos (M2)

AMB	DC	EFF	CR	COM	PRG	Easy counterfactual (used for intermediate solution)?
0	1	0	0	0	1	-
0	1	0	0	1	0	-
0	1	0	0	1	1	-
1	0	0	0	1	0	-
1	0	1	0	0	0	X
1	0	1	0	0	1	X
1	0	1	0	1	0	-
1	0	1	0	1	1	-
1	1	0	0	0	1	-
1	1	0	0	1	0	X
1	1	0	0	1	1	-
1	1	1	0	0	0	X
1	1	1	0	0	1	X
1	1	1	0	1	0	X
1	1	1	0	1	1	-
0	1	0	1	0	1	-
0	1	0	1	1	0	-
0	1	0	1	1	1	-
1	0	0	1	0	0	-
1	0	0	1	1	0	-
1	0	1	1	0	0	X
1	0	1	1	1	0	-
1	0	1	1	1	1	-
1	1	0	1	0	1	-
1	1	0	1	1	0	X
1	1	0	1	1	1	-
1	1	1	1	0	1	X
1	1	1	1	0	1	X
1	1	1	1	1	1	-

Figure A1: Necessary conditions dc + EFF and amb + PRG for positive evaluation (POS)
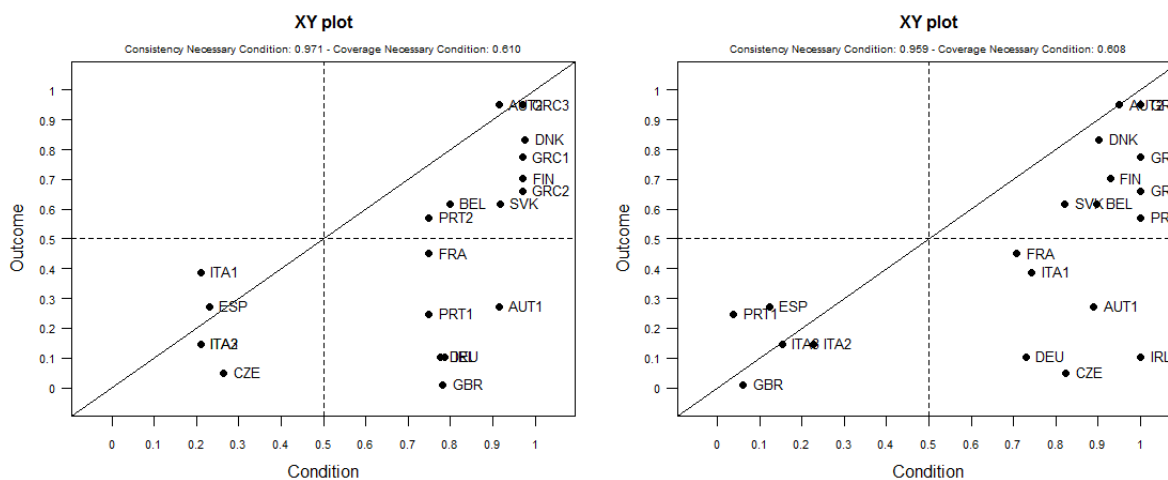

Table A9: Post-QCA case discussions

Case	Type of case	Comparison with	Question
AUT2	Most deviant case for coverage with positive evaluation	Case with similar configuration of conditions and negative evaluation: AUT1	Which additional condition distinguishes AUT2 from AUT1, fostering a positive evaluation?
AUT1	Most deviant case for coverage with negative evaluation	Case with similar configuration of conditions and positive evaluation: AUT2	Which additional condition distinguishes AUT1 from AUT2, fostering a negative evaluation?
FRA	Deviant case consistency in kind: configuration of conditions should imply positive evaluation, but did not	Cases with positive evaluation, members of the same path of the solution term DNK, FIN	Which additional condition(s) do these cases not display and fostered a negative evaluation in France?

Based on Schneider & Rohlfing 2013.

Online Appendix B: Calibration and robustness test

With the exception of the condition PRG, we used the direct method of calibration, which applies a logistic function to assign the raw data to the different qualitative categories partitioned by the qualitative anchors 0.95 (fully present), 0.5 (point of indifference) and 0.05 (fully absent) (Schneider & Wagemann 2012: 35-39). The most important anchor is the crossover point (0.5): if a change in this anchor leads to a case displaying a qualitatively different membership in the set, then this can change its membership in the truth table rows and, hence, the substantial results (Schneider & Wagemann 2012: 287-291). Conversely, changing the thresholds for full (non-)membership does not affect truth table row membership and, hence, the substantial results (Skaaning 2011). Below, we outline the calibration decisions and possible alternative crossover points. For complexity reasons, robustness tests are restricted to the complex solution. They involved the following steps for each indicator set and the outcome set (see Table B1):

1. Do the theoretical/ conceptual criteria leave room for doubt when defining the crossover point?

2. If yes, what is the *conceptually meaningful* alternative possible crossover point that still complies with the theoretical argument (Skaaning 2011: 395)?

3. Are there any empirical cases situated within the range of the old and the new crossover point (Figure B1), and if so, how many?

4. If yes: does changing the crossover point, ceteris paribus applying the calibration in Table 2, alter

   a) The cases' distribution in the condition or outcome set such that the set is so skewed that it poses severe analytical problems (see Schneider & Wagemann 2012: 232-250)? We consider this as given if the proportion of cases with membership > 0.5 is $\leq 25\%$, or $\geq 75\%$; or if the new set is much more unfavorably skewed, as compared to the original set.

b) If no: The substantial results of the analysis of necessity (in terms of a new necessary condition, or a previous one disappearing)?

c) One or several cases' membership in the truth table rows?

d) The setting of the raw consistency threshold, in terms of different truth table rows being coded as (not) sufficient for the outcome? The detailed decisions for setting the raw consistency thresholds are documented in the attached R code.

5. If yes: how does this affect the results of logical minimization?

a) Does this yield a different complex solution? For the sake of simplicity, we do not assess the robustness of the intermediate solution or parsimonious solution.

b) If yes: is the new solution term in a super- or subset relation with the original solution term? If the new complex solution term is not a subset of the original intermediate or parsimonious solution term, then the new intermediate and parsimonious solution terms will be different, too. The deviant results are reported in Table B2.

c) Which calibration scenario is to be preferred? See criteria in legend of Table B1.

Positive evaluation of austerity program by IMF. We consider the IMF's evaluation of the austerity program to be positive when the IMF considers the program to at least meet its target, which is expressed by index values of zero or more. The crossover point is set between 0 and the least negative evaluation (-0.3) at -0.15. No alternative crossover point can capture the meaning of this rating. We then coded the most extreme cases as fully in or out, respectively (full membership 1.4, full non-membership -2.4, as Great Britain (-3.7) seems to be an outlier).

Highly ambitious program (AMB). Rather than evaluating an "objective" ambitiousness of austerity programs, it can be assumed that the IMF assesses the relative ambitiousness of the programs. We hence adopt a relative perspective and code the two most ambitious programs (Portugal 1 and Ireland, which is an extreme outlier with a value of 0.6) as fully in (0.2), and the least ambitious program as fully out (0.01). For the crossover point, we choose a remarkable gap in the values, which almost double from 0.087 (FRA) to 0.15 (ITA2) (crossover point in-between at 0.1205). Alternatively, the relative perspective could be expressed by using the

sample mean as a crossover point – which also happens to be 0.12, meaning no case will be concerned.

Strongly decentralized political system (DC). Ideal-typical situations of "full" or "non-existent" autonomy are never observed in reality. Furthermore, degrees of decentralization must be interpreted relative to other countries. Accordingly, the empirical distribution of the cases, as well as in-depth case knowledge guided the choice of calibration anchors. The calibration should reflect "generally accepted notions in the social sciences" (Schneider & Wagemann 2012: 32) as to what constitutes a decentralized country. The anchor for full-membership is set at 55 to exclude Austria and include Spain, since Spain, Belgium and Germany are widely considered to be the only "real" federal countries in our sample. Countries with "intermediate" decentralization values are Czech Republic (50), UK (49), Finland (45), Denmark (42) and France (42). To establish a difference in kind between rather strong versus rather weak decentralization, the sub-index "Financial Decentralization" is arguably decisive in determining the veto power of sub-national units. The more financial issues are in the hands of sub-national units, the less room for maneuver remains for central governments to implement austerity measures. Czech Republic (rank 8) and UK (rank 11) rank considerably higher than Finland (rank 15) and France (rank 16) in terms of financial decentralization. Hence, the crossover point needs to be established at 47 - exactly between UK and Finland. Alternatively, the crossover point could be set to the sample mean (46.14) to express relatively high vs. relatively low degrees of decentralization. The anchor for full non-memberships is set at 33.5, exactly between Greece and Slovakia: according to the country profiles, Greece is much more centralized than Slovakia and Ireland.

Effective state administration (EFF). The absolute index scores have no meaning other than positioning the country with respect to other countries. Hence, the scores of individual countries can only be interpreted relative to other countries' scores. Large gaps in the cases' distribution

indicate that there is a significant (qualitative) difference in administrative efficiency between two countries. Hence, the full-membership anchor (2.065) is established between Finland (2.24) and Austria (1.89), and the anchor of full non-membership (0.7) between Slovakia (0,88) and Greece (0,52). Another significant gap between Ireland (1.31) and Portugal (1.04) separates the more effective Western European administrations from the less effective Southern and Eastern European administrations (crossover point 1.175). Alternatively, the sample mean (1.19) could serve as a crossover point to express a fully relative perspective on administrative effectiveness being above or below average.

Strong Centre-Right government (CR). The share of government-posts held by Centre-Right candidates can be considered large enough to exert significant influence on policymaking and implementation if a decisive majority of more than 50% of government posts is held by Centre-Right candidates. The crossover point is set at 52% due to a gap in the empirical values. The crossover point could alternatively be set at 49% to express the idea of a "blocking minority" rather than a decisive majority. The anchor for full membership is set at 95% to ensure that only cases in which there is not a single (potentially influential) ministerial post held by a Christian- or Social-Democrat are considered in full. The anchor for full non-membership is set at 5%, where the presumed influence of the Centre-Right on the implementation of austerity policies can be considered negligible.

High economic competitiveness (COM). The competitiveness scores represent a ranking. Individual scores do not display any objective meaning, but can only be interpreted with regard to other scores. Larger gaps imply larger ranking differences. The anchor for full membership (5.285) is set exactly between Finland/Denmark/Germany, which have top rankings, and Belgium/UK, which have good, but not exceptional rankings. We set the anchor for non-membership (4.28) in between Slovakia (which, together with Greece, ranks very low), and Italy, which is in a more intermediate cluster together with Portugal. For the crossover point

(4.86), we identify a qualitative difference between France (5.13) and Ireland (4.59). The country profiles reveal that Ireland's ranking in two aspects crucial for economic competitiveness – quality of the infrastructure and capacity for innovation – is decidedly inferior to that of France (France ranks 4 and 8; Ireland 69 and 31). Alternatively, the crossover point could be set at the sample mean (4.7) adopting a purely relative perspective. This calibration leads to a recoding of Ireland from non-competitive to competitive. It produces a solution which is identical in three of four paths and covers the same cases. The only difference is that high competitiveness contributes to a negative evaluation in two, instead of one, paths. Since this result intuitively makes much less sense, we opted for the original calibration.

Existence of an IMF program (PRG). No alternative calibration is possible.

Figure B1: Calibration and raw scores distribution
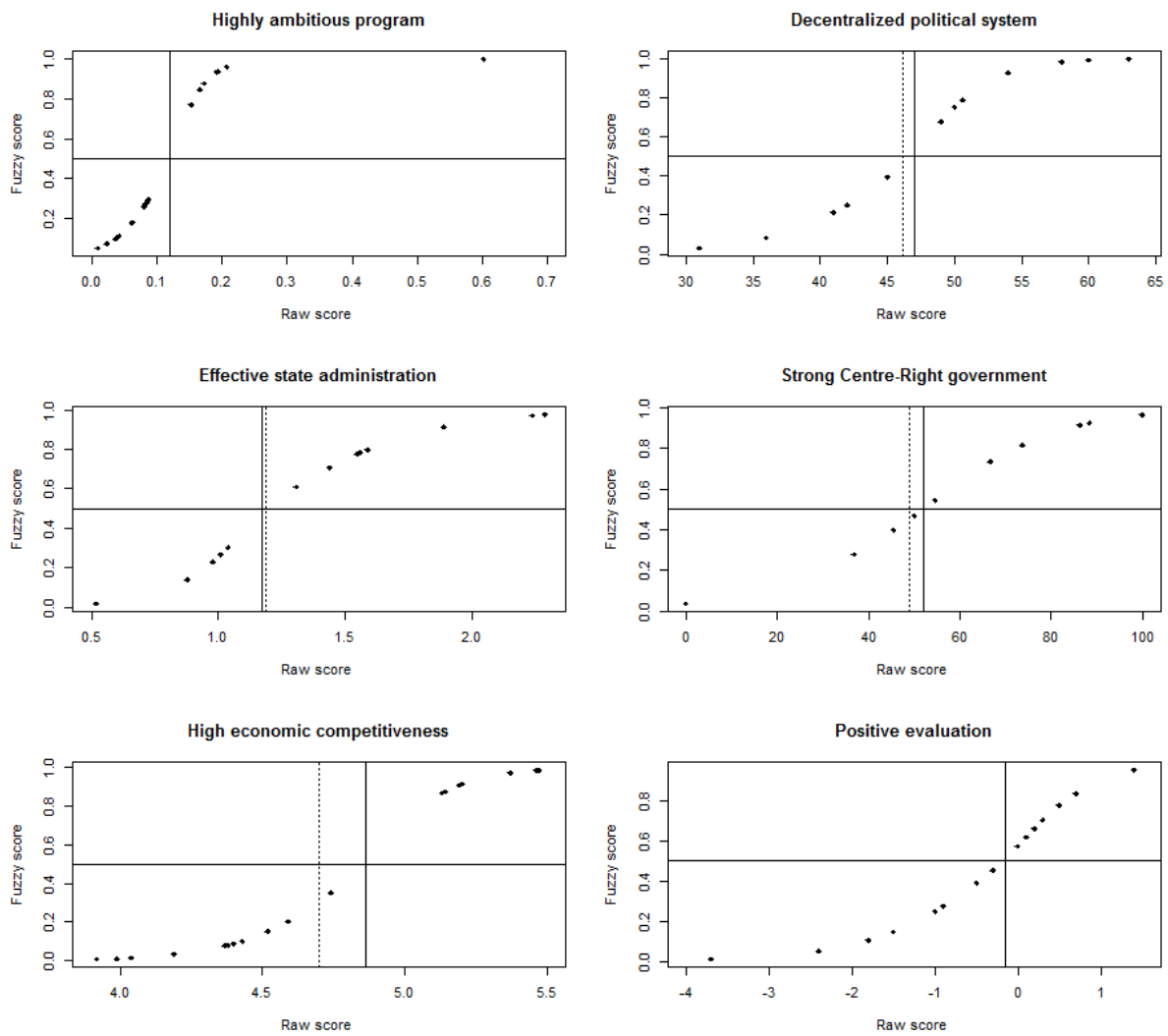
Dotted lines indicate alternative crossover points that were tested.

Table B1: Step-wise robustness check

Step	1: Alternative crossover point plausible?	2: Crossover point	3: Empirical cases within new plausible range	4b: Skewedness (%)	4c: Different results necessity?	4d: Different truth Table row membership?	4e: Different raw consistency threshold?	5a: Different solution?	5b: Super- or subset relation with old solution terms	5c: Preferred calibration
POS	NO	-0.15		45						YES
AMB	YES	0.1205		35						YES[1]
AMB2		0.12	0	35						
DC	YES	47		50						YES[1,2]
DC2		46.14	0	50						
EFF	YES	1.175		45						YES[1,2]
EFF2		1.19	0	45						
CR	YES	52		65						YES[4]
CR2		49	2	75						
COM	YES	4.86		40						YES[2,3]
COM2		4.7	1	45	NO	YES	YES	YES	NO	
PRG	NO			25						YES

[1]The non-preferred calibration does not result in a different qualitative classification of the cases and hence will not affect the substantial results.

[2]The non-preferred calibration unnecessarily contradicts recommendations of good practice (e.g. using descriptive statistics for calibration although theoretical criteria exist, or interpreting numeric values although they are not qualitatively meaningful (Schneider & Wagemann 2012).

[3]The set or the results derived from the alternative calibration have a less meaningful interpretation than the preferred set.

[4]The non-preferred calibration results in a problematically skewed set.

Table B2: Deviant results of robustness tests (sufficiency)

	Complex solution	*Consistency*	*Coverage*	*Super-/subset of old solution term*
COM2	DC*eff*com2*prg + AMB*eff*cr*com2*prg + AMB*dc*EFF* CR***COM2***PRG + AMB*DC*EFF* CR*COM2*prg → pos	0.970	0.611	No

Deviant results are marked bold.