CERRE

CENTRE ON REGULATION IN EUROPE

Independence, accountability and perceived quality of regulators

A CERRE Study

Chris Hanretty (UEA)
Pierre Larouche (Tilburg, College of Europe)
Andreas Reindl (Leuphana)

Brussels, 6 March 2012

Centre on Regulation in Europe (CERRE) asbl
Rue de l'Industrie 42 Box 16 - B-1040 Brussels - Belgium
Ph: +32 (0)2 230 83 60 - Fax: +32 (0)2 23 0 83 70 – info@cerre.eu – www.cerre.eu

CENTRE ON REGULATION IN EUROPE

CERRE CENTRE ON REGULATION IN EUROPE

Contents

About CERRE . . . 7
About the authors . . . 8
Executive summary . . . 9

1. Background **13**
1.1. The current situation . . . 13
1.2. Empirical literature on the consequences of independence . . . 18
1.3. Empirical literature on accountability . . . 19
1.4. Case selection . . . 20
1.4.1. Why these sectors? . . . 20
1.4.2. Why these countries? . . . 21
1.5. Plan of the report . . . 21

2. Independence **23**
2.1. Conceptual clarification . . . 23
2.2. Mapping the stages of independence . . . 27
2.3. Mapping the regulators . . . 32
2.4. Conclusion . . . 35

3. Accountability **37**
3.1. Conceptual clarification . . . 37
3.2. Mapping the stages of accountability . . . 40
3.2.1. Criteria where all regulators score the same . . . 40
3.2.2. Preliminary analysis: criteria with low discrimination parameters or high threshold values . . . 42
3.2.3. Remaining accountability criteria . . . 43
3.3. Mapping the regulators . . . 48
3.4. Conclusion . . . 50

4. Perceived quality **51**
4.1. Conceptual clarification . . . 51
4.2. The set-up . . . 52
4.3. The respondents . . . 54
4.4. Respondents' consistency and reliability . . . 54
4.4.1. Consistency . . . 56
4.4.2. Reliability . . . 57
4.4.3. Strategies for dealing with consistency and reliability . . . 59
4.5. Modelling perceived quality: baseline model . . . 59
4.6. Modelling perceived quality: biases . . . 62
4.6.1. Nationalism . . . 64
4.6.2. Regulator-boosting . . . 65
4.6.3. Regulatee bashing . . . 65
4.6.4. Corrected results . . . 65

4.7. Discussion of results 68

5. Empirical links between independence, accountability and perceived quality 71
5.1. Regulator resources 71
5.2. Resources, independence and accountability 74
5.3. Quality and inputs 74
5.4. Overall results 76
5.5. Conclusions 78

6. Conclusions and policy recommendations 79

A. Appendices 87
A.1. Measures of consistency and reliability for paired comparison data . . . 87
A.2. Statistical model of *de jure* independence 89
A.3. Coding of legislation 90

List of Figures

2.1. Map of provisions 28
2.2. Regulators' independence 33

3.1. Map of accountability provisions 45
3.2. Map of regulators' accountability 49

4.1. Screen shown to respondents 53
4.2. Regulator perceived quality, baseline model 61
4.3. Regulator perceived quality, alternate specifications 63
4.4. Corrected results 66
4.5. Regulator perceived quality, alternate specifications 67
4.6. Regulator perceived quality by country 69
4.7. Regulator perceived quality by sector 70

5.1. Regulators listed by resources 73
5.2. Relationship between different inputs 75

List of Tables

1.1. List of regulators 22

2.1. Independence: our working definition 26
2.2. List of items measuring independence 29

3.1. Accountability: our working definition 40
3.2. Accountability items 41
3.3. List of items measuring accountability 44

4.1. Responses 54
4.2. Pairwise comparisons 55
4.3. Kappa: rules of thumb 58
4.4. Sources of bias 64

5.1. Regression models of quality 77

A.1. Respondent consistency 88
A.2. Accountability items 90
A.3. Independence items 91

CENTRE ON REGULATION IN EUROPE

About CERRE

Providing top quality studies, training and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network industries. CERRE's members are regulatory authorities and operators in those industries as well as universities. CERRE's management team is led by Dr Bruno Liebhaberg, Professor at the Solvay Brussels School of Economics and Management, Université Libre de Bruxelles.

CERRE's added value is based on:

- its original, multidisciplinary and cross sector approach;
- the widely acknowledged academic credentials and policy experience of its team and associated staff members;
- its scientific independence and impartiality.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many member states, regulators are part of a relatively recent profession.

This study has received the financial support of CERRE members. As provided for in the association's by-laws, it has been prepared in complete academic independence. The contents and opinions expressed therefore reflect only the authors' views and in no way bind the members of CERRE.

About the authors

Prof. Pierre Larouche is Professor of Competition Law at Tilburg University, as well as a Professor at the College of Europe (Bruges). He co-founded the Tilburg Law and Economics Center (TILEC) and was, until December 2011, a founding joint academic director of CERRE. Before starting his academic career in 1996 at the University of Maastricht, he clerked at the Supreme Court of Canada in 1991-1992 and practised law for three years in Brussels. His teaching and research interests include competition law, telecommunications law, media law, basic Community law and the common European law of torts.

Dr. Chris Hanretty is Lecturer in Politics at the University of East Anglia, and member of the Centre for Competition Policy. After graduating from Oxford with a first-class degree in Politics, Philosophy & Economics, he worked for the United Nations Human Settlement Programme, and completed a PhD at the European University Institute in Florence. His research interests include institutional design and the independence of non-majoritarian institutions such as regulators, courts, and public broadcasters. He has published in the *British Journal of Political Science*, the *European Journal of Political Research*, and the *Journal of European Public Policy*.

Prof. Andreas Reindl is currently Visiting Professor at Leuphana University, Lüneberg, Germany. He previously served as the Executive Director of the Fordham Competition Law Institute, and as Principal Administrator at the OECD where he worked on competition law enforcement issues, international cooperation, and technical assistance. Prof. Reindl has also practiced U.S. and European antitrust law with the Washington, D.C. and Brussels Offices of Skadden, Arps, Slate, Meagher & Flom. Prof. Reindl has taught U.S. antitrust, European and international antitrust, and antitrust and intellectual property at Fordham Law School, the University of Michigan Law School and the American University Washington College of Law. He has published widely in the areas of antitrust and intellectual property law.

Executive summary

In this report, we investigate three key features of regulatory agencies: their independence, their accountability, and their perceived quality. We investigate these three factors for sixteen regulators drawn from four sectors (telecoms, energy, competition, and rail) operating in five different countries (Belgium, France, Germany, the Netherlands, and the United Kingdom).

Our report is based, first, on a comprehensive analysis of the legal situation regarding independence and accountability in each of these countries and sectors, and, second, sophisticated modelling of independence and accountability which relates multiple aspects of these features to a single underlying trait. Our report is the first empirical study to test the links between independence, accountability and quality across multiple sectors.

Our findings address each of these features, but also address the links between these features and the policy recommendations that flow from these findings.

Concerning independence, we find that the core of legal independence is found in thirteen different criteria, including statements of independence from market players (1) and political actors (2), clear incompatibility rules concerning positions in regulated industries (3) and in politics (4), a clear separation from the sponsoring ministry (5), exclusive and fairly detailed competences (6), a lack of any ministerial power to instruct (7) or over-rule (8) the regulator, agency budgets that are separate from the general state budget (9) and controlled by the agency itself (10), specified term lengths for heads of agencies and/or board members (11) with constraints on re-appointment (12), and limits on dismissal (13).

Concerning the measurement of independence, we find that many of these criteria can be reduced to form a single overall measure. This overall measure shows a clear progression from criteria which frequently fulfilled – such as clear statements of agency independence and limits on ministerial powers to over-rule the agency – to criteria which are fulfilled in far fewer cases, such as constraints on re-appointment and long term-lengths for members of the regulator.

Concerning accountability, we find that the core of legal accountability is found in twenty-four different criteria, which can be grouped according to the organisation(s) to which the regulator is accountable. Thus, we discuss criteria concerning accountability

towards political institutions (10 items), accountability towards the public and the market (10 items), and accountability towards judicial institutions (1 item), European authorities (2 items), and accountability towards peers (1 item).

Concerning the measurement of accountability, we find that some criteria relating to accountability – such as whether or not the objectives of the regulator are well-defined in legislation, and whether or not the regulator is accountable via horizontal networks – are present in all of the regulators we look at. We cannot therefore include them in a single overall measure of accountability. Concerning the remaining criteria relating to accountability, we find a progression from requirements to follow certain basic procedural requirements, to requirements to make public certain types of information, to requirements to periodically evaluate the regulator's work.

Concerning quality, we argue that peer evaluations of regulatory quality are a suitable proxy for the overall, or 'true' level of quality of a regulator. In particular, we argue that pairwise comparisons of regulators give consistent and reliable estimates of the quality of regulators in different sectors and in different countries. These estimates are consistent across different types of 'judges', be they academics, regulators, or regulatees. The only adjustments that must be made are adjustments to regulators who promote their own organisation.

Concerning the links between independence, accountability, and quality, we find that there is both (1) a statistically significant and positive link between independence and perceived quality, and (2) a statistically significant and positive link between accountability and perceived quality. What is more, independence and accountability are themselves positively related. This suggests that robust independence and accountability measures can effectively co-exist and contribute to better outcomes. Evaluating the joint impact of independence and accountability is difficult because of the limited number of regulators in our sample, and because of the important confounding effect of the level of resources available to the regulator. However, we present evidence suggesting that none of these links is due to coincidence.

Accordingly, our study provides empirical support for the EU approach to the design of regulatory agencies, which combines independence and accountability and seeks to find the optimal balance between them.

Concerning the policy implications stemming from this work, we suggest that (1) regulators push for greater independence and accountability in the light of the positive

effects on perceived quality; (2) that regulators seek out 'low-hanging fruit': that is, that they try to acquire criteria with low scores on our maps of independence and accountability, rather than immediately trying to acquire 'high-scoring' criteria; (3) that increases in independence and accountability be pursued together as part of balanced packages; (4) that high-scoring regulators (on our measures) focus on the breadth and quality of their involvement in horizontal networks and judicial action; and (5) that regulatees make common cause with regulators to push for greater accountability.

Acronyms used[1]

AdC Autorité de la Concurrence

ARAF Autorité de Régulation des Activités Ferroviaires

ARCEP Autorité de Régulation des Communications Électroniques et des Postes

BKartA Bundeskartellamt

BNetzA Bundesnetzagentur

CC Conseil de la Concurrence

CRE Commission de Régulation de l'Energie

CREG Commission de Régulation de l'Électricité et du Gaz

IBPT Institut Belge des Services Postaux et des Télécommunications

NMa Nederlandse Mededingingsautoriteit

Ofcom Office of Communications

Ofgem Office of Gas and Electricity Markets

OFT Office of Fair Trading

OPTA Onafhankelijke Post en Telecommunicatie Autoriteit

ORR Office of Rail Regulation

SRTF Service de Régulation du Transport Ferroviaire

[1] The full list of regulators included, together with their sectors and countries, is found in Table 1.1 on page 22.

1. Background

1.1. The current situation

European sectoral regulators can no longer be described as being in their infancy. Though patterns differ across countries and across sectors, most independent regulatory agencies are now in their teenage years, if not already in young adulthood. There are, of course, exceptions to this generalization – the German Bundeskartellamt will celebrate its fifty-fifth anniversary next year, whilst some rail regulators have only been created in the past two to three years. Yet this generalization is one of many concerning European sectoral regulators. We know broadly how and why patterns in the establishment of these regulatory agencies differ: regulatory agencies were set up earliest, and granted most independence, in political systems with relatively few veto players in which there was frequent alternation in government of opposing political parties or coalitions, and in sectors most characterised by high technical complexity and network effects (Gilardi, 2002, 2005; Elgie and McMenamin, 2005). This trend – which has now spread across Europe and indeed worldwide – has sometimes encountered resistance. For instance, national public law communities (academics, practitioners) have typically been reluctant to accept that regulatory agencies be made independent from the legislative and executive powers. Furthermore, there is continued skepticism about the degree to which formal provisions regarding the independence of these regulatory agencies translate into real independence in the day-to-day work of those agencies (Maggetti, 2007; Yesilkagit and Van Thiel, 2008). Nevertheless, those regulators which were established in Western Europe in the eighties and nineties have, by now, developed track records that enable them to be evaluated on the quality of their work.

It is precisely this issue, of the quality of regulators' work, that is addressed here. The establishment of independent regulatory agencies was justified on the basis that regulation carried out by independent agencies would deliver better outcomes than regulation by ministries. This claim was in turn based on the idea that political principals have time-inconsistent preferences, which lead them to prioritise low consumer prices over investment in infrastructure, and which generally 'spook' investors (Kydland and Prescott, 1977; Majone, 1996), and on the idea that sectoral regulation often requires a degree of technical complexity which cannot be met by an organisation headed by a politician (Majone, 1994). As a consequence, politicians – be they in the legislature or the executive power – are liable to stray from the public interest and take decisions

which are misinformed or influenced by rent-seeking on the part of market players. Yet, as we go on to argue, demonstrating this hypothesis – that more independent regulators deliver better quality performance or outcomes – has been difficult.

Overall, in Europe policy makers have implicitly or explicitly accepted the reasons developed in political science to justify the creation of independent regulatory agencies – the technical complexity of the regulated markets, the need to insulate regulation from short term political priorities, the creation of a more predictable business environment that is more conducive to investment, and, ultimately, the hope of superior market performance. But the law has not yet developed a minimum legal threshold for independent regulatory authorities that can effectively ensure that the stated goals of independence are met effectively. Thus, a "legal" concept of independence with uniformly applicable legal requirements to safeguard independence is lacking.

At least historically, the difficulty in developing a coherent legal concept of independence might have been rooted in the constitutional law traditions of European member states. Independence has always been regarded as the essential, core concept for the judiciary, but not as a feature of the executive branch where ministerial control over all administrative decisions and ministerial responsibility towards parliament are the foundational principles. Adhering to this strict dichotomy would seem to leave little room for independent decision making within the executive branch.

More specifically, in a classical separation of powers perspective, the independent regulatory agencies are caught between a rock and a hard place, conceptually.[2] If they are independent, they should belong to the judiciary power (or at least be like the judiciary), in which case they should be tasked with the mere interpretation and application of the law (according to the classical theory). Indeed, a number of authors are justifying the existence of independent regulatory agencies under a classical perspective by arguing that they are simply implementing policy decisions made by the executive (and for which the executive is accountable to the Legislature). Yet such a line of argument cannot resist a reality check. Regulatory matters cannot be split neatly between policymaking and implementation; rather, they consist in complex decision-making chains, ranging from the most general to the case-specific (Larouche and De Visser, 2006). Policy decisions are being taken through most of the chain. According to the classical perspective, then, regulatory agencies cannot both be independent from the Legislature and executive

[2] The discussion in this paragraph is sometimes presented as a Continental (as opposed to common law) perspective on independent regulatory agencies. However, in the US and the UK, the same debates took place, albeit at an earlier stage, given the longer tradition of independent agencies in these jurisdictions (since the 1930s for the US, since the 1980s for the UK).

and engage into policy-making. In practice, independent agencies are now firmly established in the European regulatory landscape, and they conduct policy-making. This led one of us, in a recent piece, to propose an alternative theoretical perspective that would fit the observed practice better (Hancher and Larouche, 2011, 743-4):

> Recent developments point towards a generalization of the conflict-of-interest rationale: in short, even if Member States have no direct interest in any of the market players, regulatory matters are high-stake games where market players will deploy considerable resources to try to influence the outcome (rent-seeking behaviour). Regulatory decisions must therefore be made in an environment which is shielded from undue influence as much as possible: this would imply transparency, independence of the decision-maker, openness, a duty to state reasons and the possibility of review, i.e. the characteristics of a regulatory agency. By implication, the role of the Legislative and the Executive would be limited to issues where there is no clear controversy among market players, i.e. issues where a decision does not immediately make winners and losers. This would explain why, in a decision chain model, the Legislative and the Executive can deal with the highest levels – provide guidelines and set out policy objectives – but cannot go very far down the decision chain, since very rapidly market players will begin to hold diverging views on the outcome and will engage in rent-seeking behaviour. [footnotes omitted]

Furthermore, even once it is accepted that regulatory agencies can be independent from the executive and legislative powers, constitutional expectations of political responsibility and control over the executive branch remain. This is why, as we explore in this report, a specific European model of regulatory agencies has emerged, which combines independence with accountability. Striking the balance between the complementary, yet potentially conflicting goals of independence and accountability is not an easy task. Solutions vary widely among member states, and sometimes among independent regulators in the same member state.

EU law adds another layer of complexity in the European model of regulation, since all independent regulators covered by this report oversee markets characterized by a direct and strong influence by EU law. It is worth emphasizing that the phenomenon of independent regulatory agencies has developed at a national level quite independently of EU law requirements. Most Member States came to recognize that independent agencies were an appropriate vehicle to address regulatory concerns in a number of

areas. EU law merely provided an additional – and sometimes decisive – impetus in favour of independent agencies. Such agencies also exist in areas which are not directly influenced by European law.

The concept of independence has a long history in European law, both with respect to regulated markets and in other contexts. In Directive 88/301 on competition in the markets in telecommunications terminal equipment [1988] OJ L 131/73, the Commission introduced at Article 6 an obligation upon Member States to entrust the regulation of terminal equipment to a body independent from market parties active in the provision of telecommunications services or equipment (in a context where Member States typically still owned the monopoly provider of telecommunications services, this was also branded as the separation of regulatory and operational functions). More than twenty years ago, the Court of Justice upheld this provision, finding that independence from market players was required by the principles of equality of opportunity between market players, effective competition and transparency.[3] Giving a market player control over, or even greater influence upon, regulatory matters would put that player in a position to distort competition. While the independence of regulatory agencies from market players has not significantly been challenged since, independence from the legislative and executive branches of government has had a more complicated story. Originally, the latter form of independence was seen as a consequence of the former: because Member States still held a significant part, if not outright control of, the largest operator, there was a risk of conflict of interest, within the government, between the interests of the State as shareholder and as regulator. Accordingly, the two interests needed to be entrusted to separate entities within the government (typically the ministries of finance and of industry, respectively).[4] This was still far from independence. Over the years, the relevant EU legislation has become increasingly detailed as regards the relationship between the regulatory authority and the executive and legislative powers. In parallel, European courts have upheld the relevant provisions, but without going as far as to derive from them a general principle of independence of regulatory authorities.[5] In

[3] ECJ, 19 March 1991, Case C-202/88, *France v. Commission* [1991] ECR I-1223 at para. 51-52. See also ECJ, 13 December 1991, Case C-18/88, *RTT v. GB-Inno-BM* [1991] ECR I-5973 at para. 25-26.

[4] See for instance Directive 90/387 on Open Network Provision [1990] OJ L 192/1, Article 5a as added by Directive 97/51 [1997] OJ L 295/23. This provision is now found in Article 3(2) of Directive 2002/21 on a common regulatory framework for electronic communications and services [2002] OJ L 108/33.

[5] See ECJ, 6 March 2008, *CMT v. Administracíon del Estado* [2008] ECR I-1265, where the Court leaves open the possibility that a ministry might be in charge of at least part of the numbering-related functions reserved to regulatory authorities under Directive 2002/21, provided that the requirements of the Directive are complied with. Taking this line to its logical end, the Court (6 October 2010, *Base*, nyr) also accepted that a Legislature might act as the regulatory authority for some of the functions listed in Directive 2002/22 on universal service [2002] OJ L 108/51, but it insisted that the Legislature must meet all requirements regarding competence, independence, impartiality, transparency and judicial

its furthest reaching judgment, the Court struck down German legislation which would have pre-empted the decision of a regulatory agency, but stopping short of formulating a general principle of independence.[6] It is only with the latest set of Directives, in 2009, that, beyond detailed provisions contributing to independence, a general principle of independence towards the legislative and executive organs is mentioned for the first time.[7]

To illustrate this point, in rail transport the focus of European law remains on the regulator's legal and functional separation from market actors; although even in this respect the exact independence criteria remain unsettled as illustrated by pending litigation against several member states. And to the extent European law requires the independence of regulators also from political institutions (such as in communications and energy markets), it does so without systematically developing a set of 'political independence criteria' that regulatory authorities must meet. Rather, as indicated above, EU law (until recently) tended to emphasize the elements of independence, in specific provisions, rather than the principle as such.

Similarly, even though, as we explain in this report, the balance between independence and accountability defines the European approach to regulatory agencies, EU law shies from explicitly setting out a general principle of accountability (which might be reliably derived from the common constitutional traditions of the Member States). Instead, EU law, in the sectors under study, is rife with provisions that can be construed as elements of a general accountability principle, be they obligations of transparency, reporting, consultation, etc., or mechanisms of judicial review, peer control through regulatory networks or control via the Commission. As with independence, the number of such elements has grown over time. In parallel, the number of principals has proliferated that can hold an independent regulator accountable, some have asked whether there may be too much accountability in the European regulatory model which could interfere with

review. Considering that no legislative body is likely to meet these requirements, the Court is therefore making it almost impossible for Member States to keep regulatory functions within the Legislature (or by extension within the executive).

[6] ECJ, 3 December 2009, Case C-424/07, *Commission v. Germany* [2009] ECR I-11431, para. 53-54. Throughout this judgment, the ECJ is careful to keep its reasoning within the scope of the provisions of Directive 2002/21 and to avoid making a general pronouncement on the independence of regulatory authorities towards the executive and legislative powers.

[7] The new electricity and gas directives are the most explicit. Directive 2009/72 concerning common rules for the internal market in electricity [2009] OJ L 211/55, compels Member States at Article 35 to make the regulatory authority "functionally independent from any other public or private entity" and give it the autonomy to decide "independently of any public body" (see Directive 2009/73 [2009] OJ L 211/94, Art. 39 for gas). In electronic communications, Directive 2009/140 [2009] OJ L 337/37 adds a provision to the Framework Directive (Article 3(3a)), stating that regulatory authorities must "act independently… from any other body", yet only as concerns *ex ante* regulation and dispute settlement tasks.

the regulator's work. Our study, however, was not designed to resolve that question.

As a result of this situation, there is a rich pool of accountability mechanisms and there is law on accountability, but an even greater lack of attempts to develop a set of uniformly applicable, legal accountability criteria.

One goal of our study has therefore been to identify sets of 'core' legal criteria, for both independence and accountability, that we could use to assess and compare the regulatory authorities covered in this report.

1.2. Empirical literature on the consequences of independence

Much has been written on the consequences of independence. In particular, independent regulators have been associated with higher interconnection rates (Edwards and Waverman, 2006) and greater firm efficiency (Li and Waddams, 2011) in telecoms markets; greater financial leverage in a range of utilities sectors (Bortolotti et al., 2011), higher capital ratios in banking (Gilardi and Servalli, 2011) and increased levels of investments in a variety of regulated sectors (Cambini and Rondi, 2011).

However, much of this literature suffers from one of three problems:

- **Poor measurement of independence**. Many scholarly articles on the independence of regulators – particularly the independence of regulators in the telecoms and electricity sectors – measure independence poorly. Specifically, many articles use a simple dichotomy – either an independent regulatory agency has been set up (and so is assumed, ipso facto, to be independent), or it has not. These articles go on to demonstrate the outcomes associated with this dichotomous measure. This dichotomous measure is unhelpful in the contemporary European experience, where the presence of independent regulatory agencies is almost uniform, but where degrees of independence differ considerably.

- **Poor choice of outcome**. Whilst many of the outcomes cited in the above literature are important in themselves, they do not tap broader aspects of the work of regulatory agencies. Financial leverage or interconnection rates are, after all, only one outcome influenced by regulators, and it may be that whilst independent regulators have recognisable effects on these outcomes, they may also have other unknown and potentially ill-advised (or beneficial) effects on other outcomes. What is needed is therefore a more global overview of the performance of regulatory

agencies.

- **Disentangling causal paths**. Because many of these studies focus on particular outcomes such as interconnection rates, they must control for a number of other factors which might also affect interconnection rates, including but not limited to, changes in the economy, existing market structure and changes in that market structure, and other barriers to entry. These 'other factors' are numerous and cannot be ignored – but the need to include factors such as this means that estimates of the effects of independence on 'quality' outcomes are dependent on the degree to which the authors have controlled for such factors. A more direct assessment of the link between independence and quality is therefore desirable.

In this report, we hope to meet these problems directly by using a sophisticated measure of independence and tying it to more direct measures of regulators' quality. Before doing so, we spend some time detailing the specific regulators that we focus on.

1.3. Empirical literature on accountability

Much has been written on accountability, but few attempts have been made to develop an empirical measurement for accountability of independent regulators and even fewer to empirically establish the consequences of accountability on performance. Accountability is a loosely defined term which means different things to different authors, including good governance, transparency, equity, integrity, openness and dialogue with citizens, and democracy. In the European literature on accountability, for example, many scholars have focused on the role of accountability in the governance structure of the European Union and on the accountability deficit (for example, Curtin et al. 2010), an approach to accountability that has limited relevance for the topic of our study. The lack of a consistent conceptual framework and incoherence in the accountability discussion has been connected with lack of empirical progress in the field of accountability standards (Bovens, 2010).

The narrower topic of accountability of regulatory authorities has of course also been addressed in the literature, which has identified various instrumental and structural mechanisms for enhancing accountability *ex ante* or *ex post* (Maggetti, 2010). In addition to the more traditional top-down mechanisms (focusing on accountability to the democratic legitimate principal) and bottom-up mechanisms (focusing on accountability to stakeholders), the literature has also highlighted the importance of 'horizontal accountability' towards peers, mostly within European networks (Hood, 1995). We were able

to use these efforts to map various accountability mechanisms in order to develop our own set of accountability criteria. There have been very few attempts in the literature to quantify accountability in a cross-sectoral and cross-country study (but see Koop 2011), and we believe that our study is one of the first attempts to examine an empirical link between accountability and quality in such a setting.

1.4. Case selection

In this report, we consider sixteen regulators altogether. These regulators are listed by country and sector in Table 1.1; the acronyms we use to refer to them are intuitive for the most part but are also listed in full on page 12.

We note that some of these regulators cover multiple sectors. The NMa, for example, has competence in the energy and rail sectors. In this report we do not explicitly consider the dual character of such regulators. We give each regulator a single score on independence, accountability, and perceived quality. We do not, therefore, assess whether these regulators perform better work when they act as, say, energy regulators rather than general competition authorities. This may be an avenue for future research.

1.4.1. Why these sectors?

These sectors were chosen for four separate reasons:

- first, because they represent a mix of sectors in which independent regulatory authorities are well-established (competition authorities) and sectors in which independent regulatory authories have been relatively recently established, if at all (rail);

- second, because whilst work has been done on the effects of independence in energy and telecoms, the same cannot be said of the effects of independence in competition (since effects are too broad) and in rail (because too new);

- third, because the sectors broadly matched the interests of CERRE members, who come from a range of sectors; and

- fourth, because including competition authorities (which are not strictly 'sectoral') makes it easier to put the perceptions of quality that we discuss in section 4 on a common scale. This is so because these authorities are found in all of the

countries we examine, have a long history upon which they can be evaluated, and are comparatively well-known.

1.4.2. Why these countries?

The choice of countries is less well-motivated than the choice of sectors. This is because the choice of countries largely reflected the interests of CERRE members. Nevertheless, the five countries in question display a variety of legal systems and variety in the timing of their establishment of independent regulatory agencies, from very early adopters (the United Kingdom) to relatively late adopters (Belgium).

1.5. Plan of the report

Our report proceeds through each of the features we identify as key – independence, accountability, and quality. In chapter 2, we discuss the key features of independence, and in particular the way in which we can model independence as consisting of a series of items located at different 'thresholds' separating lower- from higher-independence regulators. In chapter 3 we proceed in a similar fashion, discussing the elements of accountability, and how this too can be modelled. In chapter 4 we discuss measures of quality in regulation, and how a suitable proxy for quality can be found in the peer evaluations of regulators compared two-at-a-time. Finally, in chapter 5 we use the measures of independence and accountability discussed in chapters 2 and 3, together with the measures of perceived quality presented in 4, to tease out the links between these three aspects.

Table 1.1.: **List of regulators**

	Competition	Telecoms	Energy	Rail
Belgium	Conseil de la concurrence	Institut belge des services postaux et des télécommunications	Commission de Régulation de l'Électricité et du Gaz	Service de Régulation du Transport Ferroviaire
France	Autorité de la Concurrence	Autorité de régulation des communications électroniques et des postes	Commission de Régulation de l'Energie	Autorité de Régulation des Activités Ferroviaires
Germany	Bundeskartellamt	Bundesnetzagentur	[Bundesnetzagentur]	[Bundesnetzagentur]
Netherlands	Nederlandse Mededingingsautoriteit	Onafhankelijke Post en Telecommunicatie Autoriteit	[Nederlandse Mededingingsautoriteit]	[Nederlandse Mededingingsautoriteit]
United Kingdom	Office of Fair Trading	Office of Communications	Office of Gas and Electricity Markets	Office of Rail Regulation

2. Independence

In this section we discuss the concept and the measurement of independence. We begin by noting some conceptual problems common in discussions of independence. We then move on to measurement of independence in terms of a select number of items, or provisions, which reflect higher or lower levels of independence. We map these items and discuss their contribution to independence. We then move on to showing how the regulators in our study score on this measure of independence.

2.1. Conceptual clarification

Our hypothesis is that levels of independence affect regulators' perceived quality. In other words, more independent regulators do better work. However, there are a number of difficulties involved in measuring independence which must be discussed before we can talk about regulators having more or less independence.

The first issue we must deal with lies in specifying *what kind of independence*, and *independence from whom*. It is common to distinguish between *de jure* and *de facto* independence, or the independence that an organisation has according to the law, and the independence that an organisation has according to practice. Here, we concentrate on *de jure* independence. We do so for three reason. First, dealing with *de facto* independence involves very difficult judgement calls; there is the risk that any measure of *de facto* independence might either be too subjective, or too broad, to be satisfactory. Second, dealing with *de jure* independence has clear policy recommendations – for whilst it is possible to recommend that the law be changed, and relatively easy to change statutes, it is much harder to recommend that practices be changed, and harder still to change those practices. Third, although we concentrate on *de jure* independence, we know that political science studies assessing both *de jure* and *de facto* independence have found that these two aspects are correlated (Hanretty and Koop, 2010). In fact, as the legal requirements concerning independence are specified in greater detail (going into resources, appointments, etc.), the room for *de facto* independence to diverge considerably from *de jure* independence is reduced (save for the unlikely case of a wholesale failure to respect legislation). Thus, any improvements to *de jure* independence may ultimately result in improvements to *de facto* independence.

As indicated in Section 1, it is also important to distinguish between independence from

the executive and legislative powers (i.e. independence from politics) and independence from regulated interests. The latter is less controversial (in both theory and practice) and as such it will not be considered further in this report. The majority of our items used in measuring independence concern independence from politics. We concentrate on independence from politics because, as set out above, much of the literature on the creation of independent regulatory agencies has argued that these agencies have been created because elected politicians take poor (temporally-inconsistent, non-Pareto-optimizing) decisions, and that agencies can take better decisions if they are independent of these politicians. Although we do not discuss independence from regulated interests, later we discuss extensively accountability towards regulated interests through the publication of relevant information and other rules on disclosure.

The law on independence also reflects these two dimensions, i.e., independence from market participants and independence from political institutions, although the latter is the more recent feature. Historically, as mentioned in Section 1, the focus has been on independence from market players (i.e. the separation of regulatory and operational functions), which figured in early European legislation and which the Court rapidly established as a core element in regulated markets. The Commission and the Court derived the requirement to separate regulators and regulated market players from competition law principles, thus implicitly assuming that market performance would improve with greater independence of regulators.

Until recently, only independence from market participants was specifically required in European legislation.[8] At the same time, the EU seemed to take a piecemeal approach to independence from politics: whilst it was not mentioned, a number of elements supportive of independence from politics were required by EU law, so that Member States were gently nudged towards accepting. Indeed, it could be observed that, at Member States, there was a greater, yet uneven, willingness to make regulatory agencies independent from politics as well. A prime example can be found in national competition authorities, to which Member States have granted independence from politics as if it was a matter of course, despite the absence of any requirement to that effect in Regulation 1/2003 or its predecessor.

As noted in Section 1, independence from the executive and legislative powers has only recently been mentioned in directives concerning energy and communications markets. Here as well, beyond a brief general clause, EU law continues along its piecemeal ap-

[8] Today independence from market participants remains the sole legal requirement in EU legislation concerning rail transport, although here, again, most member states are willing to go beyond the European requirements.

proach, with the addition of new specific elements conducive to independence, apparently in response to events in certain Member States that could be seen as an attack on the independence of regulators. The reactive nature of the legislation may also explain the appearance of ad-hoc measures that regulate certain independence criteria in great detail, but lack provisions related to other criteria. For example, the 2009 revisions to the electronic communications framework directive add detailed provisions concerning the dismissal of heads of agencies, but lack anything on their appointment as well the length of the term and its renewability.

The concept of independence from political institutions may be quite recent for regulated sectors, but it has a long tradition in other areas of European law. The understanding of what was legally required to ensure independence can to some extent enrich the discussion of what constitutes the 'core' legal concept of independence for market regulators.

By way of illustration, independence has been a key element in distinguishing 'courts or tribunals' that may refer questions on European law to the Court of Justice under Article 267 TFEU from regular administrative agencies that do not have the right to refer. Even if the analogy has its limits, elements in the case law on what constitutes an 'independent' tribunal are relevant also to establish independence of regulatory authorities, such as the Court's requirements of specific and limited reasons for termination and dismissal, and of a prohibition against ministerial instructions or supervision. Vague dismissal provisions, on the other hand, have been identified as a risk to complete independence.

In a case examining limitations to the independence of the European Central Bank, the Court identified several provisions at the core of independence, such as the prohibition against any influence by other Community institutions or member state governments, the prohibition on the ECB from seeking instructions; the ECB's legal personality, own resources and budget authority; and specific dismissal provisions. At the same time, however, the Court confirmed that certain types of control of independent authorities remain permissible if they do not undermine independent decision making in the areas which have been allocated to the authority's exclusive jurisdiction. A more recent case on the independence of data protection agencies has also clarified which hierarchical relations between political institutions and independent regulators impermissibly interfere with independence.

While the law on independence to date remains a patchwork, a Commission staff pa-

Table 2.1.: **Independence: our working definition**

Independence	A regulator is *de facto* independent to the degree that the regulator takes day-to-day decisions without receiving and acting on the basis of instructions, threats, or other inducement from politicians or market players, or the anticipation thereof; or considering whether the interests of those politicians or particular market players would be harmed by particular choices about those decisions. A regulator is *de jure* independent to the degree that the legislation or statute governing the regulator works so as to make instructions, threats, or other inducements impossible.

Note: adapted from the definition found in Hanretty (2010)

per on the independence of energy regulators (European Commission, 2010) may lay out how a more comprehensive, systematic legal approach to independence might look like in the future. Among other elements, it refer to separate budget allocation and independent spending authority; separate premises; sole responsibility for management, organization, and staffing matters; sanctions (including possibly criminal sanctions) in case of violations of independence provisions; and sufficient resources and salaries that attract qualified staff. At this point, the paper remains more like a vision, as it is unclear whether the Court will accept the paper's expansive interpretation of existing legislation, but it may identify where existing independence requirements might be tightened in the future.

The second issue we must deal with lies in specifying *what items to include in any index of independence*, and *how to weigh those items*. Independence, as found in legislation, takes many different forms. One statute might wish to guarantee the independence of an agency by making it very difficult for politicians to dismiss the head of the agency. Another statute might wish to guarantee the independence of an agency by giving it a funding stream which cannot easily be altered by politicians – such as contributions from regulated operators. It is not clear, in the abstract, whether the former should have a greater weight, or count for more, than the later. Nor is it clear whether all possible items are equally good candidates for inclusion in any eventual index.

A third issue concerns the different possibilities within each item. Some common items – such as funding, or obligations towards the executive – have multiple possibilities. An agency may face no reporting obligations to the executive (or the legislature), or may simply have to send an annual written report, or may face an annual in-person briefing at the sponsoring department. The order of these items, in terms of their con-

sequences for independence, is not, we would suggest, in doubt. Agencies which must brief their sponsoring department in person presumably face less independence than agencies which face no reporting requirement. But how big is the gap between these different options? If in-person reporting diminishes independence by x units compared to annual written reporting, does annual written reporting diminish independence by the *same amount* compared to no reporting, *more*, or *less*? These questions of spacing are difficult to tackle.

In developing our list of independence criteria, we built upon a list of criteria commonly used in the political science literature, which we then cross-checked against those criteria that most commonly appear in European legislation, case law, and legal literature on independence. Thus, we eschewed a 'state of the art' approach where we would have included every measure of independence that we could find in legal sources, and instead created a list of what we saw as 'core' independence criteria in light of their frequent use in a variety of legal sources.

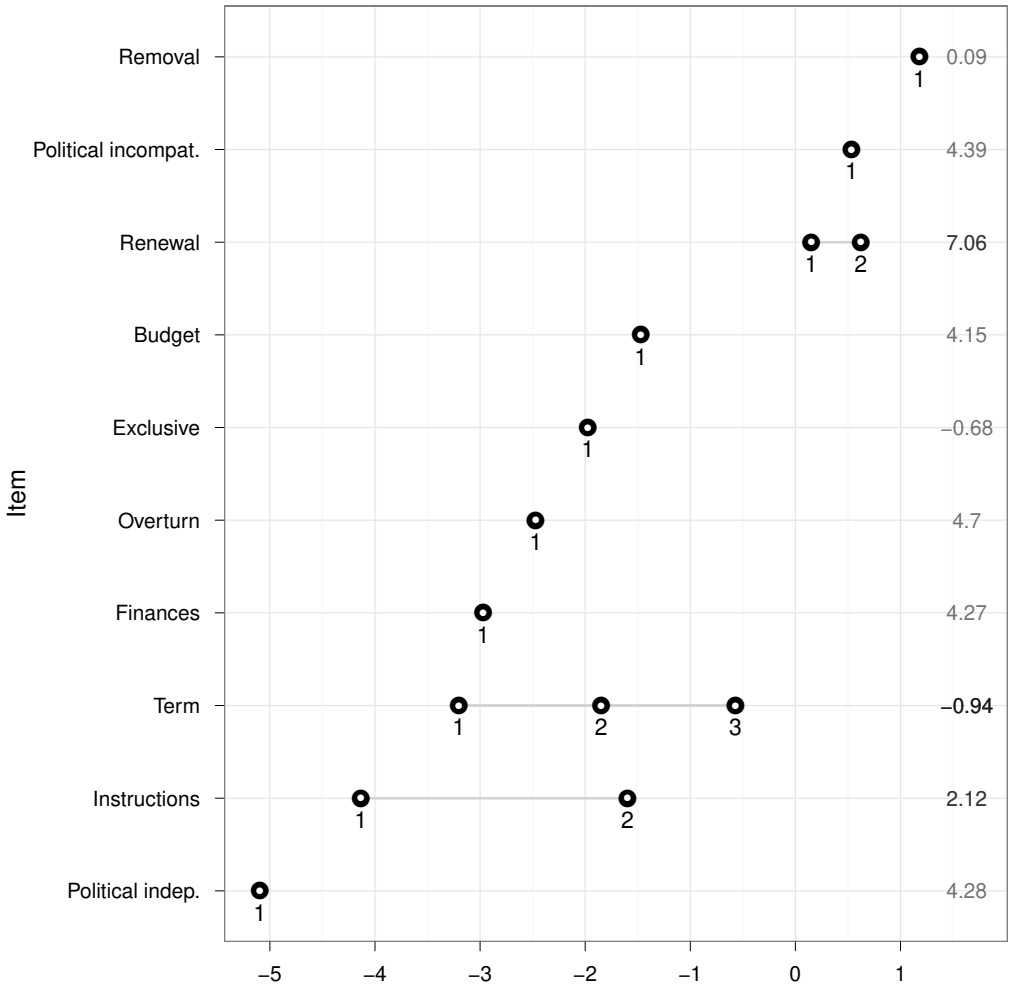
Our approach is at once simple and sophisticated. It is simple in that we allow the data to answer many of the issues we posed in terms of inclusion, exclusion, weighting, and spacing. Institutional features that are commonly found together are taken to be independence-promoting; institutional features that are commonly absent together are taken to be independence-weakening. If we have enough data available, we can calculate numbers which reflect the weighting of each item. Items that cluster with other independence-promoting items receive a high weight. Items which sometimes cluster with other independence-promoting items, but which sometimes are found alongside independence-weakening features, are given a lower weight, because they are noisier signals of the level of an agency's *de jure* independence.

The technique that we use – an ordinal factor analytic model – is described in a technical appendix. Fortunately, (almost) all of the parameters estimated by the model can be presented in an easy to interpret visual format. We list all of the items that we include in Table A.3.

2.2. Mapping the stages of independence

Figure 2.1 plots the parameters estimated by the model. Items are ranged down the left hand column, and are listed in a particular order. That is, they are listed in order of the lowest *threshold* associated with each item.

Figure 2.1.: **Map of provisions**



Note: Plotted positions refer to the threshold where two adjacent options are equally likely. Thus, plot points marked '1' refer to values of independence for which either the zero-th or the first listed option is equally likely. See the body of the text for a listing of the different options. The number plotted in the right-hand side of the map is the discrimination parameter. See text for a description of the discrimination parameter.

Table 2.2.: **List of items measuring independence**

Item	Description
Political indep.	Whether or not the regulator is described in legislation as being independent
Instructions	Whether or not the minister may instruct the regulator in (a) no issues, (b) general issues or (c) specific issues
Term	Whether the head of the agency is appointed for (a) 4 (b) 5 (c) 6 years
Finances	Whether or not agency finances are separate from state finances
Overturn	Whether or not politicians can overturn agency decisions
Exclusive	Whether or not the agency has exclusive regulatory competence in its sector
Budget	Whether or not the agency controls its own budget
Renewal	Whether the appointment of the agency head may be renewed (a) multiple times, (b) once only or (c) not at all
Political incompat.	Whether or not agency members may hold political office
Removal	Whether or not the agency head may be removed from office (other than through non-policy related issues such as sickness or incapacity)

The threshold for each item – or more specifically, for each item response – is the point at which a regulator is equally likely to have one of two adjacent item responses. Take, for example, the first threshold listed, for political independence. Here, there are two item responses:

0 No, the regulator is not described as politically independent

1 Yes, the regulator is described as politically independent

The threshold plotted is the threshold between the zero-th and the first listed option. Thus, a regulator with an independence score of approximately -5.1 would be equally likely to be not described as politically independent as it is to be described as politically independent. (Note that we cannot move the other way: just because a regulator is not described as politically independent does not mean that it has an independence score of close to -5.1. Other features of the regulator's governance might affect its score. Only if the regulator had low-independence responses for all other items could we conclude that that the regulators' independence was either roughly equal to or below). To the right of this threshold, regulators are more likely to be described as politically independent;

to the left of the threshold, they are less likely to be described as such.

This is the first listed threshold, or, what is equivalent the threshold with the lowest value. If we were speaking from the perspective of those drafting government legislation delegating power to independent regulators, we might also describe this item not just as the item with the lowest threshold, but the 'easiest' item – in this case, the easiest concession to the independence of the regulator. Alternately, if we were speaking from the perspective of a would-be independent regulator, we would describe this item as the 'easiest' to achieve.

The next 'easiest' item to achieve for a regulator that wishes to be independent concerns the instructions that may be given to the regulator. Here, however, the relevant item has two thresholds rather than one. That is because there are three possible responses to this item:

0 The minister may instruct the regulator on all matters

1 The minister may instruct the regulator, but only concerning broad frameworks outside of any individual case

2 The minister may not instruct the regulator

The easiest threshold bridges the zeroth and first responses. That is, after obtaining a statement of its independence from politics, the next 'easiest' guarantee of independence for a regulator to achieve is a practical and easily-stated constraint on the minister's action: namely, a bar on his/her ability to issue instructions on specific cases. Thus, a regulator with an independence score of approximately -4.13 would be equally likely to be able to be instructed on all issues as it would be to be instructed on framework issues only. Regulators with higher independence scores, however, would be less likely to be able to be instructed on all issues, and more likely only to be instructed on framework – up to a certain point, at which the regulator is decreasingly likely to be able to instructed on framework issues, and increasingly likely to be free from any kind of instruction.

The next item concerns the term of members of the regulator. In discussing this item, we must also discuss the *discrimination parameter*. The discrimination parameter is listed in the right hand side of the plot. It can be interpreted as the degree to which the given item successfully discriminates between regulators of similar independence. The higher the discrimination parameter, the more quickly the probabilities of answer-

ing in adjacent categories increase or decrease: in other words, the item becomes a more reliable and clear cut guide to the true level of independence of the regulator. Here, however, the discrimination parameter is negative, suggesting that the criteria is related to independence, but in the opposite direction to that predicted. Negative discrimination parameters imply that higher-ordered categories are associated with less independence rather than more (for items with multiple possible responses), or that having a certain property is associated with less independence (for items which regulators either possess or lack).

In this specific case, the negative discrimination parameter of -0.94 implies that longer term lengths of six years or more actually imply less independence than terms of four or five years. This finding does not tally with previous research on the measurement of independence (Hanretty and Koop, 2011). It is likely to be a consequence of the fact that the regulators in our sample have heads who are generally appointed for longer five- or six-year terms, rather than shorter three- or four-year terms found in other regions of the world. In this specific situation, it may mean that regulators who wish to achieve greater *de jure* independence would be better pushing for greater restrictions on politicians' ability to instruct them – or clearer separation of their finances, which is dealt with by the next item.

The item on finances is, once again, an item with only two responses. The lower independence response involves the finances of the regulator being dealt with as part of the general finances of the state. The higher independence response involves the finances of the regulator being dealt with by the regulator itself.

This item is similarly located to the item on the power of politicians to overturn agency decisions. Both items require a score of between -3 and -2 to answer in the 'higher-independence' category. In both instances the item is relatively discriminating when it comes to identifying higher- and lower-independence regulators.

The same cannot be said of the next item, which concerns exclusive competencies. Here, the negative discrimination parameter signals that having exclusive competence over a particular area is actually associated with lower levels of independence. It seems that having exclusive control over a particular area does not mean that such control will be exercised independently. This, then, is good news for sectoral regulators who operate alongside other sectoral or cross-cutting regulators. This finding also tallies with previous research on independence (Hanretty and Koop, 2011).

We skip over the item relating to control of the agency's budget (an important, and highly-located item) to briefly discuss the last three items. Whether appointments to the regulator may be renewed or not is a very highly discriminating item which is located towards the top end of the independence spectrum. The logic here is that by preventing reappointment, or by preventing reappointment more than once, regulators cannot be 'put on trial' by politicians, and only re-appointed conditional on how they have discharged their duties in previous years. This turns out to be slightly more discriminating than a clear statement of political incompatibility. In turn, both of these items turn out to be much more discriminating than the issue of removal, which strangely turns out to discriminate relatively little between regulators of otherwise similar levels of independence.

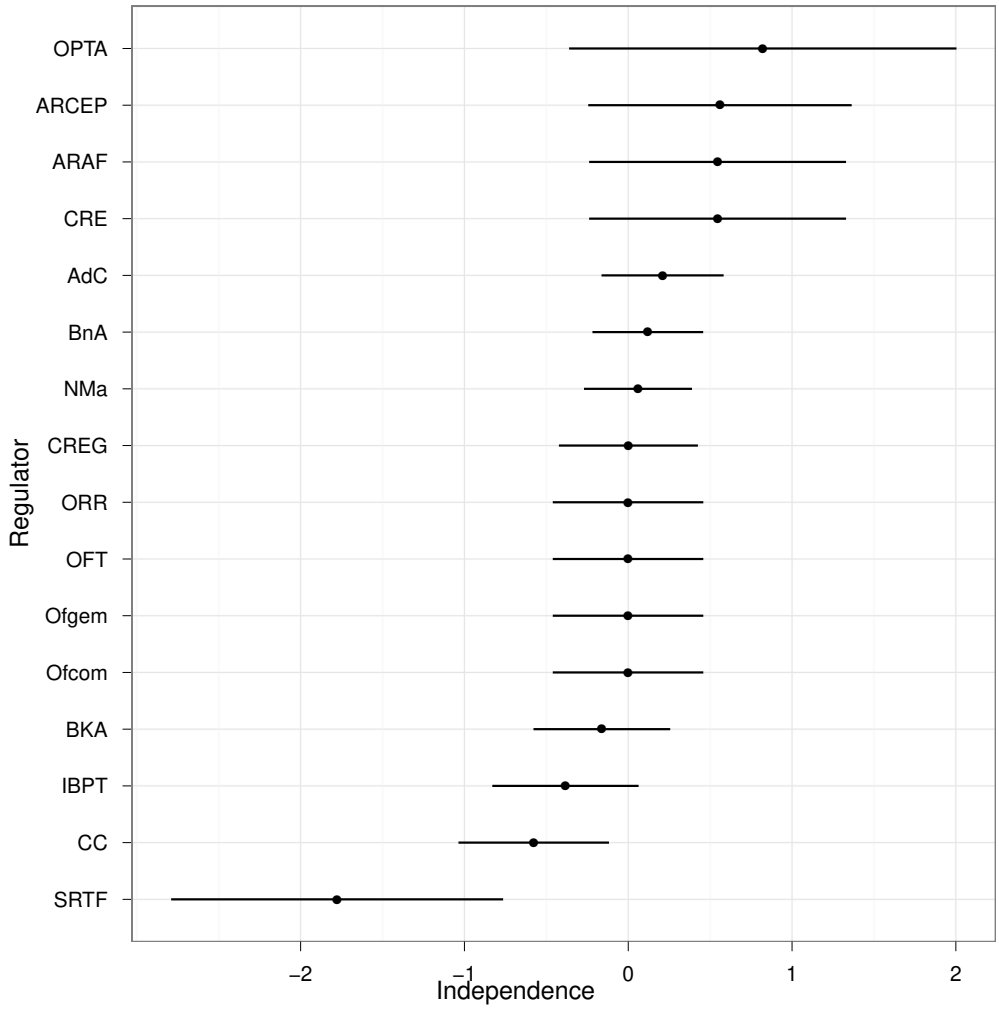
2.3. Mapping the regulators

The same model that we used to find out the relative ordering of items relating to independence can also be used to produce estimates of the independence of each regulator. These estimates are plotted in Figure 2.2. The estimates are presented with their associated 95% confidence intervals.

There is much less variation concerning the independence of regulators than there is concerning the contribution of individual items relating to independence. This is perhaps understandable. Regulators in the same country may work from essentially the same governance template, adapted only to their special needs. Consequently, we might not expect them to differ very much in their degree of independence. This means that although we can be confident that the 'true' independence of OPTA is greater than the true independence of the SRTF, to take the two regulators at opposite ends of the scale, we can be much less confident when comparing regulators in the middle of the figure. That is, we cannot be very confident that the true independence of the Bundesnetzagentur is greater than the true independence of the NMa – though this is still more likely than not. As is common with models of this kind, estimates of extreme positions are less precise than estimates of positions towards the middle of the scale.

Note that all the estimates of independence are in the range [-2,2]. This is very different from the estimates of the thresholds, which were in the range [-6,2]. This shows that we cannot easily read across from a particular feature of a regulator's governance to its independence score. In reality all regulators have governance structures that are a mix of 'high-independence' features and 'low-independence' features.

Figure 2.2.: **Regulators' independence**



Note: Dots indicate the best model-based estimate of the regulator's independence. Lines surrounding the dots show the 95% confidence interval of this estimate.

The independence scores for certain regulators are lower than we had expected based on our reading of the legislation. This is particularly the case for the two German regulators. Readers should note that the independence score of the regulators depends on an interpretation of the governance provisions in the applicable national laws. Which interpretation is preferred will affect the ranking in particular for regulators in the middle of the figure. This can be best illustrated with the example of the Bundeskartellamt. The low independence score of the Bundeskartellamt is explained primarily by the appointment and dismissal provisions concerning the head of the authority: the President's term is not defined (therefore there is no need to state anything about renewability) and there are no provisions limiting the Minister's discretion to remove the President from office. These features are, in fact, extremely rare among the regulators we have examined in this report. In addition, the minister has the power to issue general guidelines.

Yet the Bundeskartellamt has a unique governance structure: the President is head of authority, but has no influence in the decision making process. All decisions are adopted by decision making departments whose members belong to the Bundeskartellamt staff and enjoy all job protection guarantees of German civil servants. They are not subject to any instructions by the President when adopting decisions and, thus, act like quasi judicial, independent panels.

In our evaluation of the Bundekartellamt's independence, we decided to focus on provisions governing the position of the head of authority, to ensure consistency and facilitate comparability across all regulators. This is also consistent with the approach taken elsewhere in the literature using a comparative approach to examine the independence of regulators (de Visser, 2009). We accept, however, that there would be a greater degree of independence if we had focused on the level below the President. Future work on independence might address this problem with a more refined model to measure independence.

Taking the independence scores as given in the table, and noting the particular problems relating to the interpretation of the legislation governing the two German regulators, we note that the disparities in independence between sectors are less marked than the disparities between country. More specifically, whilst there is a statistically significant difference between the perceived quality of regulators according to country (F-value of 3.94 on 4 degrees of freedom, p = 0.03), the same cannot be said of quality by sector (F-value of 0.84 on 3 degrees of freedom, p = 0.5). Belgium (mean score of -0.68) and Germany (mean score of -0.02) are the countries with the least independent regulators; Netherlands (mean score of 0.44) and France (mean score of 0.47) are the countries

with the most independent regulators.

2.4. Conclusion

In this section we have outlined our approach to measuring independence, and have stated that we measure the *de jure* independence of regulators from politics, rather than actual independence, or independence considered more broadly. We have described how, rather than assigning arbitrary weights and thresholds to certain items, we have 'let the data do the work', and estimated the contribution to independence of a variety of items. We have then discussed, by way of a discussion of each item, how regulators might move from low-independence to high-independence, and have presented estimates of the level of independence of each regulator.

This measurement is essential if we are to gauge the contribution of independence to regulators' perceived quality. Only by measuring independence accurately can we test the link between these two variables. Only by measuring independence in this way can we chart the next policy-relevant steps for regulators to improve their level of independence, and thus achieve greater levels of *de facto* independence – and, as we show in section 4, greater quality.

We note the technical details of the model we used in an appendix present at the end of the report. For the moment, we turn to assessing the levels of accountability of the several regulators in our sample.

CENTRE ON REGULATION IN EUROPE

3. Accountability

Just as with the independence of regulators, the accountability of regulators needs to be clarified conceptually. Like independence, accountability as a concept benefits from a considerable 'halo effect', in that it is presumed to be positive, and measures taken to improve or extend accountability are seen positively in virtue of the positive association which attaches to the root concept. This halo, however, may obscure as well as illuminate. In the first part of this section, we discuss the concept of accountability, and relate it to the items included in a particular index of accountability. We go on to discuss how we can map the stages of accountability, and place regulators on that same metric.

3.1. Conceptual clarification

We would make three conceptual points about accountability (all of which are borrowed from Philp 2009). The first point is that accountability is a relationship between two actors. That is, accountability exists when some A is accountable to some B.

In our case, it is clear that it is the regulator which is accountable to other actors. It is not so clear to whom the regulator is accountable. Are regulators accountable to politicians (and if so, which sets of politicians: those in the national government, those in the national legislature, or politicians at other levels, such as the European level)? Or are regulators instead accountable to a broader set of actors, including but not limited to national publics and regulated actors? Here, we distinguish a number of different accountability relationships: accountability to politicians, accountability to the market, accountability to the judiciary,[9] and accountability towards relevant peer groups such as networks of sectoral regulators, or the European Commission. Some of these accountability relationships, such as accountability to the judiciary and accountability to peer groups, are tapped only by a few criteria. Some other relationships, such as accountability towards politicians and the market, are broader, and this breadth is reflected in a larger number of criteria relating to this relationship.

Our second point is that accountability is a relationship which involves 'giving account': that is, informing, explaining, and justifying conduct. Because of this, many of the items

[9] One could argue that, since the judiciary does not generally have the right to act on its own motion, accountability to the judiciary is in fact a proxy for accountability to regulated firms and – to the extent that they have standing, see ECJ, 21 February 2008, Case C-426/05, *Tele2 Telecommunication* [2008] ECR I-685 – members of the general public.

in our index of accountability relate to the provision of information: the submission of annual plans and reports, both to political institutions and to the wider public; the scheduling of face-to-face meetings with relevant parliamentary committees or advisory bodies drawn from the public or relevant market players; and publication of reasons and data underlying decisions taken by regulators.

Our third point is that accountability need not, *as a matter of definition*, involve sanctioning behaviour. Philp (2009, 35) explains the distinction as follows:

> The intuition behind the desire to make sanctioning analytically part of the definition of accountability is that, without any sanctions, accountability could be an entirely paper exercise. The intuition behind resisting this move is that A is accountable if he or she can be required to give an account, irrespective of whether certain consequences may follow from doing so. The former intuition is strongly encouraged by principal-agent thinking – since Y wants to be able to get A to do certain things and needs to use incentives and disincentives to ensure that he or she does – and by those who are concerned that without sanctions the process may be toothless. But the latter intuition (resisting sanctions) is related to the recognition that much accountability concerns imparting information, transparency, reporting and justification, and that these processes do not have to be driven by the threat of sanctions to have value.

This conceptual point is ultimately born out by our empirical analysis. Although we begin by including an index item which relates to sanctioning – specifically the power of the minister to dismiss the head of the agency – we find that this item is unrelated to the other items which measure accountability towards political institutions and other stakeholders. We consequently remove it from our index (see below).

This conceptual point also matters when we consider the relationship between accountability and independence. Accountability and independence are sometimes taken to be inimical to each other. This is based on the view that accountability requires the possibility of sanctioning, and mechanisms which allow sanctioning compromise independence towards political institutions and other stakeholders. If, however, accountability can be achieved by transparency and publication requirements alone, then the two concepts will be easier to reconcile in practice.

As regards accountability toward the European Commission and the judiciary, sanc-

tioning mechanisms exist (and, at least in the case of courts, must exist). This could be seen as a challenge to the position we adopt in this report that sanctioning behavior is not a necessary component of effective accountability mechanisms. We believe that we have a good answer to this challenge as far as the European Commission is concerned; as far as courts are concerned, we find it more difficult to develop a good answer that satisfies all three authors to the same extent, although we note that this question does not affect the results of this report.

As to accountability towards the European Commission, we accept that elaborate schemes have been developed in the areas of competition, communications, and energy that allow the Commission to influence (if not outright veto) decisions by national regulators. These could be seen as a sanctioning mechanism that is an instrumental part in the Commission's toolbox to hold national regulators accountable. We consider more significant, however, that while the threat of 'sanctions' always exists in these areas, the Commission has used its power rarely (and not at all in some areas). Communication and cooperation mechanisms shape accountability toward the Commission to a much greater extent than the 'hard instrument' of interfering with national authority decisions. Thus, transparency and publication/communication mechanisms appear to explain quite well how accountability toward the Commission works in practice.

The very role of courts in a regulatory system is to sanction. They are there to review the work of regulators, and to quash their decisions if they have made substantive or procedural mistakes. That makes it more challenging to maintain the view that accountability systems can work without sanctioning mechanism. Of course, the role of courts might develop along the same lines just described for the Commission: there are some appeals, the regulator learns its lessons from appeal decisions, and as it internalizes these experiences and adjusts its decision making practice, the regulator is increasingly held accountable through the threat of judicial review and less through actual court decisions. While this model might look quite appealing, we realize that reality looks quite different in most countries: Some regulators continue to lose a high percentage of their cases on appeal, and if anything it would appear that litigating decisions of regulators has become the standard. We find it difficult to develop a completely satisfactory answer. On the one hand, the model of accountability through internalizing mechanisms and the threat of judicial review should be examined more carefully in order to explore whether judicial accountability can be reconciled with their view that accountability is best explained by transparency and publication mechanisms. On the other hand, it could also be that courts hold regulators accountable through their ability to sanction, which in practice they also do quite frequently, even if that means that the view that accountability works

Table 3.1.: **Accountability: our working definition**

Accountability	A regulator is accountable with respect to its past actions and future intentions when national or European politicians, or market players, or members of the public, or judicial authorities, can require the regulator to inform and to explain and/or justify its conduct with respect to those actions and intentions.

Note: based on the definition found in Philp (2009)

without sanctioning mechanisms does not extend all the way to courts. We do note that while these questions are conceptually interesting and important, they do not affect the empirical work and the results of our study.

3.2. Mapping the stages of accountability

Our initial selection of accountability criteria was complicated by the fact that there appears to be less agreement in the political science literature on which criteria should be considered when determining the degree of a regulator's accountability, and a much of a less systematic treatment of accountability in legal sources. In creating our list of accountability criteria, we sought to include criteria for each of the five main groups of principals: political institutions; other stakeholders/the public; courts; European institutions; and peers. For each group, we identified criteria that are regularly mentioned in political science and legal sources, although it was much more difficult here to discriminate among criteria in order to include only those that could be identified as 'core' elements. These difficulties explain why our initial list of accountability criteria was much longer than our initial independence criteria list.

The following table provides an overview of the accountability criteria we found, together with an indication of the principal(s) which immediately benefit from it. The table regroups the criteria into four stages, which are introduced further below. It also already provides an indication of those criteria which, while relevant, were not used in the assessment, for various reasons (as explained below).

3.2.1. Criteria where all regulators score the same

The technique used in this section is similar to the technique used earlier to measure levels of independence and the contribution of each item to overall independence. Because this technique exploits variation between the responses of different regulators, it

Table 3.2.: **Accountability items**

Principal	Executive	Legislature	Commission	National public	Regulated firms	Other regulators (peers)	Judiciary
Accountability criterion							
Procedural stage							
Presence of defined regulatory objectives	X	X	X	X	X		X
Objectives explained to stakeholders	X	X	X	X	X		X
Reasoned decisions	X	X	X	X	X		X
Procedural rules in place				X	X		X
*Advisory body with stakeholder participation				X	X		
Information stage							
Report to Commission	X						
Retrospective annual report	X	X	X	X	X		
Press releases	X	X	X	X	X		
Voluntary information provision	X	X					
Discovery stage							
Compulsory/compelled information provision	X	X					
Appearance before parliamentary Committee		X					
Public consultations				X	X		
Annual reports publicly available				X	X		
Data underlying decisions publicly available				X	X		
Evaluation stage							
Prospective annual plan	X	X					
Code of conduct				X	X		
Periodic performance evaluation	X	X					
Possibility of Commission intervention			X				
Appeal before judiciary					X		X
Membership in peer network						X	
**Prior ministerial approval for annual plan	X						
**Minister can issue guidelines to regulator	X						
**Minister can dismiss head of regulator	X						

* Not used in the assessment, since all regulators score the same.

** Not used in the assessment, because of high threshold value.

*** Not used in the assessment, because of low discrimination parameter (low influence on accountability score).

is unable to 'score' items where all regulators in the sample have the same response. Three items in our index – whether the objectives of regulatory policies are defined, with priorities, whether or not the decision-making of the regulator was transparent, and whether or not the regulator belongs to a horizontal network – do not have any variation. Therefore, on the basis of the empirical data, we cannot say how much having defined regulatory objectives, or membership of a horizontal network, contributes to accountability.

The fact that we cannot 'score' these items does not mean that they are not important for accountability. Indeed, it may be that these items are fundamental for accountability. Having defined regulatory objectives might be such a basic element of accountability that it is a characteristic of all regulators in our sample, even those which score poorly on other characteristics.

3.2.2. Preliminary analysis: criteria with low discrimination parameters or high threshold values

Having excluded these three items because of lack of variation, we conduct a preliminary analysis of the data using a two-parameter item response model. This model is explained in full in the technical appendix. Just as before, we estimated *thresholds* for each item, and a number of *discrimination parameters*. These thresholds and discrimination parameters can be interpreted in the same way as before. That is, the threshold for a given item is the point on the scale at which the regulator is as likely to respond in a given high-accountability category as it is to respond in the low-accountability category. Here, the interpretation of the thresholds is made easier because we are dealing with items which either feature or do not feature; regulators which score close to the threshold are as likely to have a given accountability-promoting feature as they are to lack that feature. The discrimination parameter measures roughly how important having this item is to the regulator's overall score for accountability. High discrimination parameters indicate items that are particularly important for accountability.

When we conducted this preliminary analysis, we found that three items had very small or negative discrimination parameters. The items affected were:

- whether or not the minister may reject the regulator's annual plan or budget;
- whether or not the minister may issue general guidelines concerning the regulator's work; and

- whether the minister can dismiss the head of the regulatory for performance-related reasons.

Negative discrimination parameters sometimes arise when conducting this kind of analysis. They may mean either that the item is unrelated to the latent trait we are trying to measure, or that although the item is related to the latent trait we are trying to measure, it is related in precisely the opposite direction to that which we predicted. Given that there are no strong reasons for suspecting that items are related in the opposite way to that predicted, we view the former possibility as more likely. It is also worth noting that two of these items, concerning ministerial instructions and the possibility of dismissal of the head of the regulator, also feature in our index of independence. Our finding that these two items are unrelated to accountability but negatively related to independence may indicate that indeed, the type of measures where independence and accountability would theoretically be in direct conflict are in fact not necessary for accountability to be ensured.

Furthermore, we also leave aside one item pertaining to whether or not the regulator had an advisory body. The threshold for this item, at 19.565, was extremely high, and including it in the plot made it difficult to display the range of the remaining values. As is discussed below, the high threshold associated with this item is related to the extremely low discrimination parameter, suggesting that whilst this item is an indicator of high accountability, it is a very unreliable one.

3.2.3. Remaining accountability criteria

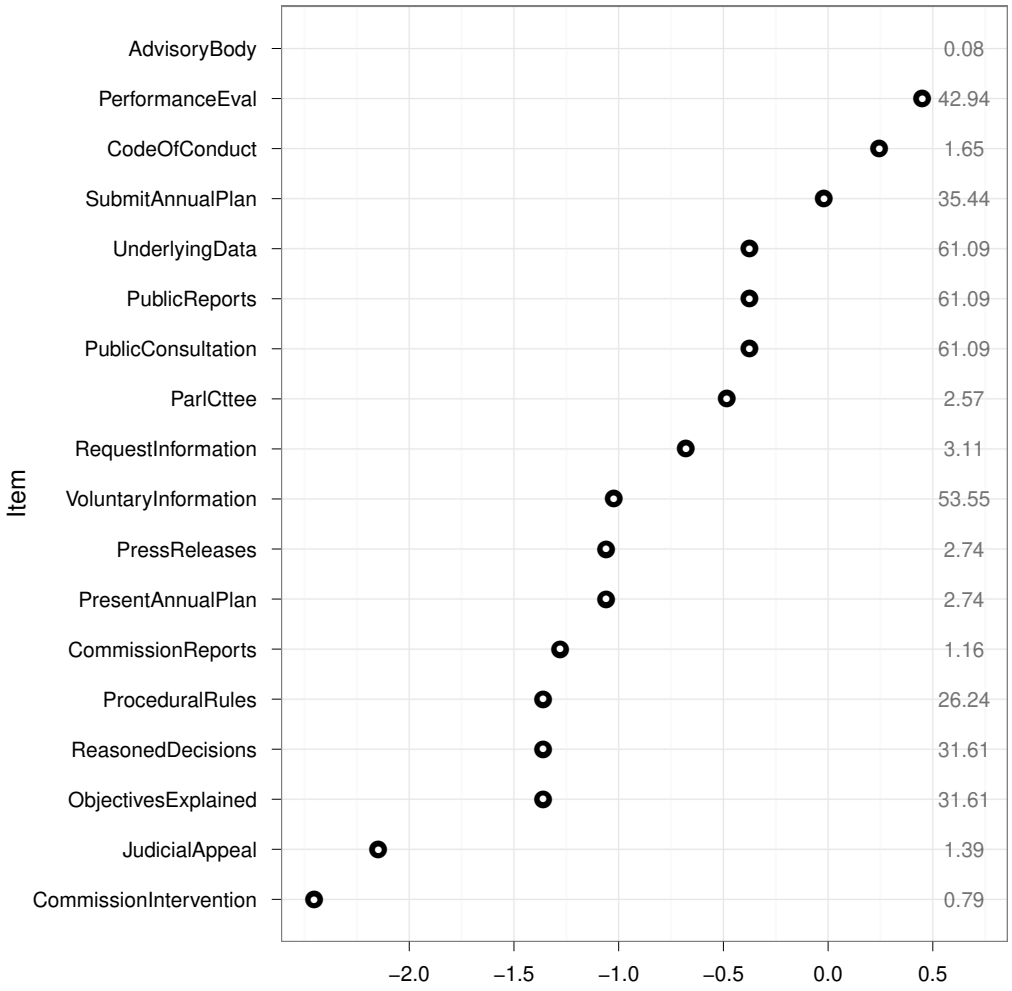
Thus, having excluded three criteria because of lack of relationship to the overall trait of accountability, and one because of a too high threshold value, we were left with seventeen items relating to accountability. These items, and their associated thresholds and discrimination parameters, are presented in more detail in Figure 3.1.

We begin by discussing the bottom two items shown in Figure 3.1, namely whether or not the Commission can intervene before decisions are adopted or finalized (*Commission Intervention*) and whether or not judicial appeals with full review are permitted (*Judicial Appeal*). Both of these items have extremely low thresholds, which means that ordinarily we would expect that even regulators which lack other mechanisms of providing accountability would have these items. Having said this, the discrimination parameters attached to these items are amongst the lowest in our index of seventeen items, suggesting that these items may not be as essential to accountability as other

Table 3.3.: **List of items measuring accountability**

Item	Description
Commission Intervention	Whether or not the commission can intervene before decisions are adopted/finalized
Judicial Appeal	Whether or not judicial appeals with full review of decisions are possible
Objectives Explained	Whether or not strategies and policy goals explained to stakeholders
Reasoned Decisions	Whether or not reasoned decisions are published
Procedural Rules	Whether or not there are public rules on procedures and participation in decisions and rule making
Commission Reports	Whether or not the regulator reports to European Commission
Present Annual Report	Whether or not the regulator presents (retrospective) annual reports and financial statements
Press releases	Whether or not the regulator issues press releases and newsletters reporting on its work
Voluntary Information	Whether or not the regulator may provide information to a minister, or the parliament, voluntarily at its own discretion
Request Information	Whether or not the regulator must provide information to a minister, or the parliament, upon request
Parl. Committee	Whether or not the regulator must periodically appear before a parliamentary committee to answer questions
Public Consultation	Whether or not the agency conducts public consultations with feedback opportunities
Public Reports	Whether or not the agency makes its annual reports publicly available
Underlying Data	Whether or not the agency makes publicly available the data underlying its decisions
Submit Annual Plan	Whether or not a (prospective) annual plan and budget are submitted to political authorities
Code of conduct	Whether or not the agency has a code of conduct/ guidelines
Performance Eval.	Whether or not the agency is subject to periodic performance evaluation
Advisory Body	Whether or not the agency has an advisory body with stake holder participation

Figure 3.1.: **Map of accountability provisions**



Note: Plotted positions refer to the threshold where two adjacent options are equally likely. The number plotted in the right-hand side of the map is the discrimination parameter. See text for a description of the discrimination parameter. Note also that the threshold for Advisory Body is not plotted. See text for explanation.

items. As indicated before, if transparency and disclosure mechanisms function well, it may be that formal sanctions (by the Commission and by the judiciary) are less frequently applied and thus become less significant in ensuring accountability.[10]

The next set of items largely concerns procedural rules concerning the operation of the regulator, and in particular the way in which the operator explains itself to regulatees and to the Commission. These four items – whether or not strategies and policy goals are explained to stakeholders (*Objectives Explained*), whether or not reasoned decisions are published (*Reasoned Decisions*), whether or not there are rules on procedures and participation in decisions and rule making (*Procedural Rules*), and whether or not the regulator reports to the European Commission (*Commission Reports*) – all have relatively low thresholds. What is more, with the exception of the item relating to reporting to the European Commission, they all discriminate fairly well between regulators with similar levels of accountability. The lower discriminatory ability of the item relating to reporting to the European Commission is likely to reflect a difference between sectors rather than a real difference between comparatively accountable and comparatively unaccountable regulators. We would therefore suggest that these four items constitute the first major milestone on the road to accountability, and should be treated as a package.

The next three items all concern the ways in which the regulator communicates information to the outside world. All three items concern basic ways in which the regulator may be held accountable by other actors (politicians, stakeholders, the broader public) acting on information released by the regulator. Whether or not the regulator presents its annual reports and financial statements to political institutions (*Present Annual Plan*) and whether or not the regulator issues press releases/newsletters to interested parties (*Press Releases*) are two basic ways in which the regulator communicates; the regulator may also go beyond this, and provide information to parliament and to the minister according to a reporting schedule of its own devising (*Voluntary Information*), a step which although it is similarly demanding in terms of its threshold, is ultimately more discriminating than either of the two provisions concerning mandatory reporting and communications activity. Nevertheless, these three items pertaining to communication follow basic provisions concerning procedure, and form a second package relating to accountability mechanisms.

These two packages, concerning procedural rules and communications respectively,

[10] We suggest in our conclusion that what may matter is not the brute fact of participation in judicial processes, but rather the quality of such participation.

are followed by another less tightly clustered package concerning the information upon which the regulator bases its decisions – which perforce includes a strong communications element. The first of these items concerns whether or not the sponsoring minister and/or the parliament may request information of the regulator (*Request Information*). Like the next item along, concerning whether or not the regulator is obliged to appear before relevant parliamentary committees and answer questions (*Parl Cttee*), this item discriminates reasonably well – but not as well as the next three items, which concern, respectively, whether or not the regulator consults the public (*Public Consultation*), whether or not the annual reports of the regulator are made publicly available and readily accessible (*Public Reports*), and whether or not the data underlying the regulator's decision is also made public (*Underlying Data*). These three items have the highest discrimination parameters of any items in our index, and effectively separate high- from low-accountability regulators.

The remaining items in our index concern different ways in which regulators may set standards for themselves upon which basis they will later be evaluated. The first of these items concerns whether or not the regulator submits a (prospective) annual plan of work, as distinct from a (retrospective) annual report. This item discriminates well, especially in comparison to the next item, concerning whether or not the regulator follows a code of conduct, or guidelines for its operations (*Code Of Conduct*), which, although more demanding, is less discriminating than the previous item. The last item which reliably discriminates between regulators concerns whether or not the regulator is subject to periodic performance evaluation (*Performance Eval*). This is the most demanding item, with the exception of whether or not the regulator has an advisory body. This last item has a very high threshold, but has such a low discrimination parameter that its impact on overall scores for accountability is negligible, and indeed calls into question the relevance of this item to the overall index.

This brief discussion of eighteen index items related to accountability, the 'difficulty' of these items, and the degree to which these items discriminate between regulators with otherwise-similar levels of accountability, strongly suggests that accountability proceeds in four schematic stages:

- a *procedural stage*, in which the regulator must state the rules on the basis of which it is to be held accountable, and demonstrate that it is following these rules;

- an *information stage*, in which the stakeholders – both political actors and the wider public – acquire the information to hold the regulator accountable for its

actions;

- a *discovery stage*, in which stakeholders – again including political actors and the public – may, through the examination of select information, examine particular actions of the regulator; and

- an *evaluation stage*, in which the regulator states the objectives on the basis of which it is to be held accountable, and is subject to evaluation on this basis.

Interestingly, this four-stage schematic overview of accountability is without sanctioning elements. Political scientists have debated whether or not the prospect of being sanctioned by a given actor is an inherent part of being accountable to that same actor. As announced above, this index suggests that accountability can exist largely on the basis of disclosure of information alone. That may provide an explanation why, as we discuss later, there is a positive relationship between accountability and independence.

3.3. Mapping the regulators

Just as before with our model of independence, the same model that we used to find out the relative ordering of items relating to accountability can also be used to produce estimates of the accountability of each regulator. Once again, we present these estimates alongside indications of the uncertainty surrounding each estimate. Whilst there is greater separation between regulators in terms of accountability than there is in terms of independence, the results must still be interpreted cautiously. These estimates of accountability are provided in Figure 3.2. As before, 95% confidence intervals are also plotted.

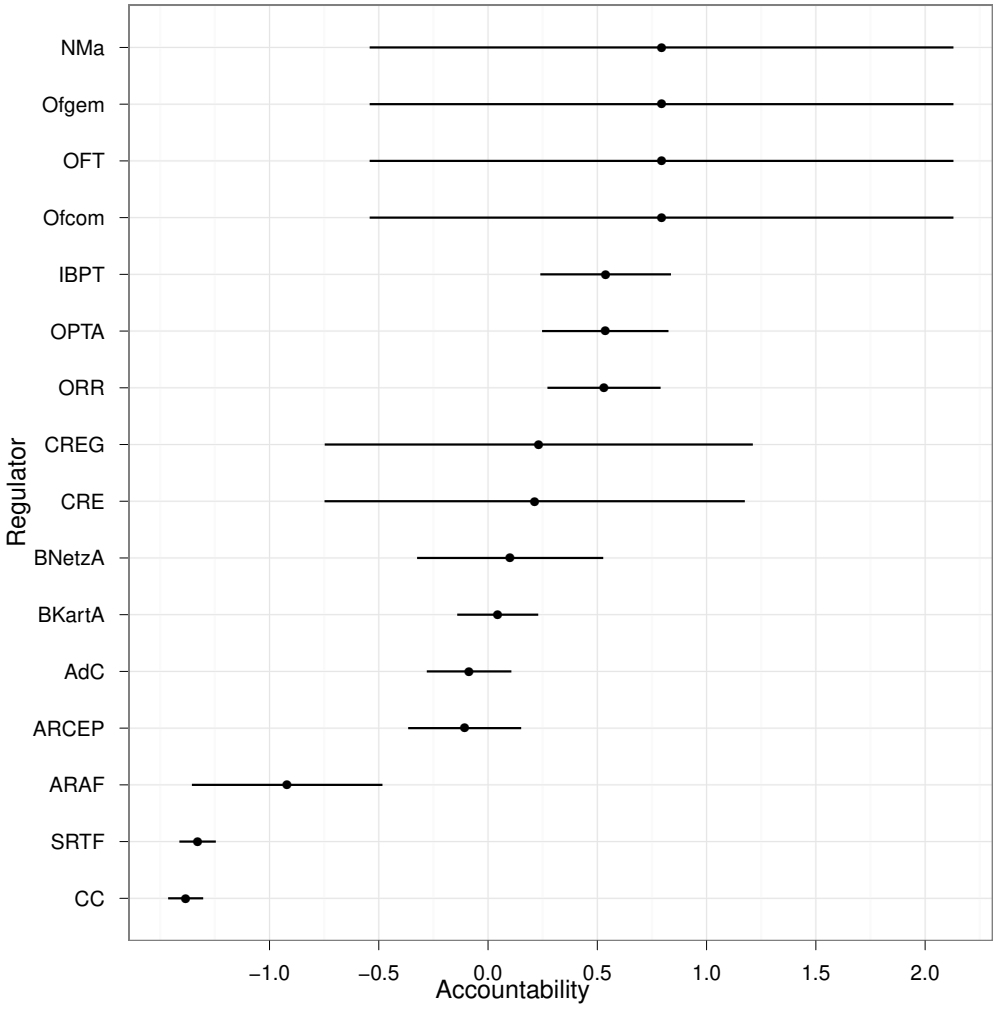
The estimates of accountability grow increasingly imprecise as we move from low accountability to high accountability.

Of the three regulators with the lowest levels of accountability, two – ARAF and SRTF – deal with the regulation of railways. We can be highly confident that these regulators have lower levels of accountability than the other regulators in the sample. Both agencies are quite young, which may partly explain why they have less developed accountability mechanisms and score lower on our map.[11]

Though we cannot be equally confident, it is probable that the next group of regulators

[11] ARAF, for example, has not yet published an annual report because it has been so recently established.

Figure 3.2.: **Map of regulators' accountability**



– the two French regulators in telecoms and competition respectively, and the German Bundeskartellamt and Bundesnetzagentur – are in the bottom half of the sample when it comes to accountability. The accountability of this group of four regulators is surrounded by comparatively little uncertainty.

The same cannot be said for the two energy regulators, the CRE and the CREG. Our estimates of accountability for these two regulators are subject to considerable uncertainty.

The top seven regulators fall into two groups of regulators, each impossible to judge between. The first group is composed of the ORR, the Dutch OPTA, and the Belgian IBPT. The second group is composed of the three remaining British regulators, and the Dutch competition authority.

3.4. Conclusion

In this section we have discussed the concept of accountability, and how it might be measured. Setting out with an initial stock of some twenty-three items, we noted that some items were not appropriate to include in an index of accountability comparing the regulators in our sample. We then estimated a model which calculated the contribution of each of these items to regulators' accountability. We suggested that these items could be grouped into four different stages of accountability: a *procedural* stage, an *information* stage, a *discovery* stage, and an *evaluation* stage.

We concluded by presenting the accountability scores of the sixteen regulators in our sample, noting where relevant the difficulties in reliably distinguishing between regulators with similar levels of accountability.

4. Perceived quality

This section describes the results of a peer benchmarking exercise carried out at workshops organised by the Centre for Regulation in Europe (CERRE) on 29 September. The object of the exercise was to elicit attendees' perceptions of the quality of regulators in four sectors – competition, energy, rail and telecoms – in five countries – Belgium, France, Germany, the Netherlands, and the United Kingdom. The section starts by describing how attendees' perceptions were elicited, before giving some descriptive information concerning the responses and their consistency. The section then goes on to describe how the perceptions of the attendees can be turned into an overall ranking of perceived quality of the regulators, and how this ranking can be adjusted for biases that respondents may have. The section concludes by showing how perceived quality varies according to country and sector.

4.1. Conceptual clarification

As Radaelli (2004) noted,

> The concept of quality has now become a fundamental component of regulatory reform and regulatory management in a large number of countries. 'Quality' features in all initiatives for better regulation launched by the OECD and the European Union. EU institutions such as the Council and the European Commission have opted for a basic definition of 'high quality regulation'. Efficient, effective, coherent, and simple (that is, easy to understand) regulation is high quality regulation – [as] the Commission argues in its official publications … and even in its own tenders.

However, the fact that quality is much discussed does not mean that quality is well defined. As Radaelli goes on to note, quality is associated with a variety of characteristics. 'High quality' regulatory work is, variously, work that is efficient (both in its own production and in its consequences for regulated markets), proportionate, legitimate, consistent, not unduly prescriptive, and enforceable. That is, quality is made up of many potentially rivalrous components.

In this report, we do not specify the precise number and weighting of all of these components of quality. Whilst such a work might be possible, we do not believe that it is

appropriate here. Instead, we rely on a notion of perceived quality. Specifically, regulator X has greater perceived quality than regulator Y if competent experts choose X over Y when asked which regulator does better work. In relying on perceived quality, we short-circuit the question of the different components of quality, and allow our competent experts to silently weight these different components when they make their comparison.

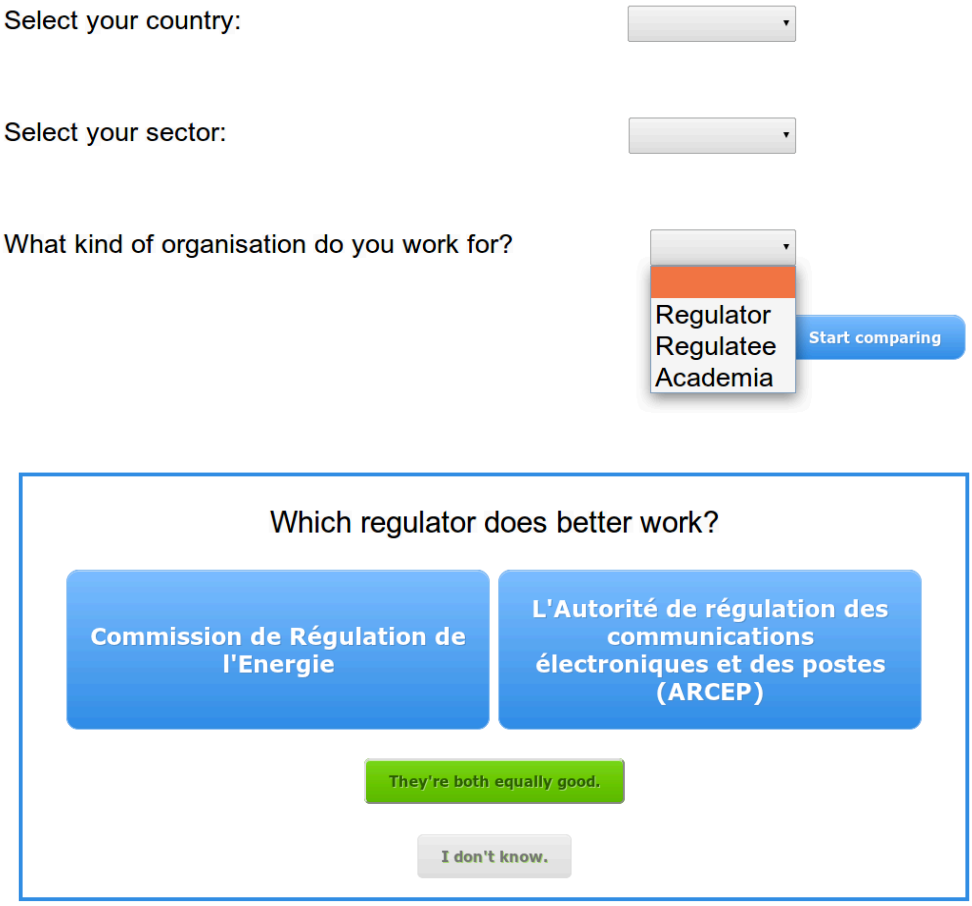
Because we rely on perceived quality, it is possible that our judgements of quality might not be convergent. In particular, experts judging the quality of two regulators might make judgements in part based on their own positions and interests. For example: regulated actors might place great weight on the efficiency and prescriptive character of 'quality regulatory work'. Regulators, on the other hand, might place great weight on the coherence and effectiveness of 'quality regulatory work'. Academics studying regulation, finally, might place great weight on the proportionality and legitimacy of 'quality regulatory work'. In other words, relying on perceptions of quality might mean that our measures are purely subjective, rather than inter-subjective.

In what follows, we try and account for this possibility, and discuss perceptions of quality across the three groups of experts we asked: regulators, regulatees, and academics. We show that responses are consistent across groups. We then go on to discuss the particular model that we use to combine expert judgements of the quality of regulators. We then present estimates of perceived quality for each regulator – the result of our combining expert judgements. Before this, we briefly discuss the manner in which expert judgements were solicited.

4.2. The set-up

Respondents were asked to take an electronic survey. They first had to indicate their country, their sector, and their role – whether regulator, regulatee, or academic. Respondents were then asked to compare every pairwise combination of regulators in their country, and in their sector. Thus, regulatees from the Belgian telecoms sector were asked to compare all pairwise combinations of the Belgian regulators (Institut Belge des Services Postaux et des Télécommunications (IBPT), Commission de Régulation de l'Électricité et du Gaz (CREG), Conseil de la Concurrence (CC), Service de Régulation du Transport Ferroviaire (SRTF)), and all pairwise combinations of the telecoms regulators in the five countries (IBPT, Autorité de Régulation des Communications Électroniques et des Postes (ARCEP), Bundesnetzagentur (BNetzA), Onafhankelijke Post en Telecommunicatie Autoriteit (OPTA), Office of Communications (Ofcom)). Respon-

Figure 4.1.: **Screen shown to respondents**



dents from the Belgian telecoms sector would therefore be asked to make sixteen different pairwise comparisons.

This second group of comparisons between regulators operating in different countries was necessary in order that all of the regulators in the sample could be placed on a comparable scale. Otherwise, we would have had a series of national league-tables, with no indication of how these national figures corresponded to one another.

An example of the prompt given to respondents is shown in Figure 4.1. The question given to respondents – Which regulator does better work? – was deliberately short and provocative. Respondents had four options in responding to this prompt. They could indicate that one of the two regulators shown did the better work; that both regulators were equally good; or that they did not have enough information to judge.

The total number of comparisons made by each respondent thus depended on the number of regulators in their country and upon the number of regulators in their sector. Belgian or British respondents could expect to make 6 of the possible pairwise combinations between the four British/Belgian regulators, and 10 of the possible pairwise combinations between the different regulators in their sector. Dutch or German respondents could expect to make a single pairwise comparison within their country, and 10 of the possible pairwise combinations between the different regulators in their sector.

4.3. The respondents

Table 4.1 summarizes a number of basic details regarding the number of responses and the number of comparisons. 'Valid responses' are simply those responses where the respondent chose one regulator over another, or said both regulators were equally good. 'Don't know' responses are thus 'invalid'.

Table 4.1.: **Responses**

Respondents	34
Potential comparisons	524
Valid comparisons	399
Average valid comparisons per respondent	11.7
% valid comparisons per respondent	76.2
(Min, max)	(18.8, 100)

In terms of the composition of respondents, the most popular sectors were Rail (11 respondents), Telecoms (10 respondents), and Energy (7 respondents).

The most represented countries were Belgium (15 respondents), United Kingdom (9 respondents), and France (6 respondents). The ratio between academics, regulatees, and regulators was 7:16:11.

We can see more clearly the number of comparisons between the different regulators in Table 4.2, which shows the number of pairwise contests between each regulator. The table is symmetric, so the entry in row i, column j is equal to the entry in row j, column i.

4.4. Respondents' consistency and reliability

Consistency and reliability are two key issues when eliciting judgements from multiple participants in an evaluative exercise such as this. By *consistency*, we mean the ability

Table 4.2.: **Pairwise comparisons**

	OFT	ARCEP	OPTA	Ofcom	BNetzA	Ofgem	ARAF	CRE	AdC	ORR	NMa	IBPT	CREG	CC	BKartA	SRTF
OFT				9		9			6	9	6			6	6	
ARCEP			10	10	10		6	6	6			10				
OPTA		10		10	10						3	10				
Ofcom	9	10	10		10	9				9		10				
BNetzA		10	10	10		7	11	7		11	18	10	7		1	11
Ofgem	9			9	7			7		9	7		7			
ARAF		6			11			6	6	11	11					11
CRE		6			7	7	6		6		7		7			
AdC	6	6					6	6			6			6	6	
ORR	9			9	11	9	11				11					11
NMa	6		3		18	7	11	7	6	11			7	6	6	11
IBPT		10	10	10	10								15	15		15
CREG					7	7		7			7	15		15		15
CC	6								6		6	15	15		6	15
BKartA	6				1				6		6			6		
SRTF					11		11			11	11	15	15	15		

of respondents to rank regulators such that, if they prefer A to B, and B to C, then they also prefer A to C. By *reliability*, we mean the ability of respondents to give similar judgements when presented with the same comparison. If respondents are inconsistent and/or unreliable, and if this inconsistency or lack of reliability is not the result of systematic factors which can be accounted for, then our overall results concerning perceived regulatory quality may be called into question.

Accordingly, in this section we present diagnostic statistics on the consistency and reliability of respondents. We use these statistics later when we present our overall results on perceived regulatory quality. Specifically, we show that neither (a) excluding unreliable respondents nor (b) positively weighting consistent respondents affect our overall results.

4.4.1. Consistency

Our working assumption in this exercise is that participants at CERRE meetings are highly knowledgeable about regulatory activity in their country and in their sector. They ought, therefore, to be able to compare regulators and make informed comparisons between them. Sometimes, however, respondents may have knowledge of sectoral regulators in only one or two countries; or they may not have knowledge of certain regulators in their own country. In this situation, respondents may either answer truthfully that they do not know enough to compare the two regulators (an option taken in 23.9% of comparisons), or they may guess which regulator is better. In extreme cases, respondents may answer questions entirely at random. This would hardly deliver credible results.

In order to test for uninformed responses, we can examine the consistency of respondents' pairwise comparisons. If respondents are choosing regulators at random, or if they know very little about the regulators they are being asked to choose between, then we would expect to see what are known as *circular triads*. That is, if we have three regulators, A, B, C, and respondents have strict preferences (that is, they either prefer A to B: A $\succ$ B or B $\succ$ A), then we might see patterns of the form

$$A \succ B \succ C \succ A$$

demonstrating inconsistency.

In order to quantify the degree of (in)consistency, we may make use of Kendall's coefficient of consistence ζ (Kendall and Smith, 1940). The calculation of Kendall's ζ, and in particular the way in which we accounted for the possibility of ties and the presence of missing data, are described fully in the appendix. These ζ values we calculate can be used to test whether inconsistencies in the response patterns are there by chance, or whether we can conclude that the respondent was choosing regulators at random.

Possible values of ζ are between zero and one. Higher values of ζ indicate greater consistency. The average value of ζ, 0.57, indicates moderate to high consistency. Although they are not reported here, statistical tests on these ζ values showed that the null hypothesis could not be rejected for any respondent. That is, we cannot conclude that respondents were choosing regulators at random. Consequently, there are no good grounds for excluding respondents on the basis of inconsistency.

Although we cannot exclude respondents on the basis of inconsistency, we may still be interested in relative levels of (in)consistency. Analysis of Table A.1 shows that

- Academics are on average more consistent than regulatees, who in turn are more consistent than regulators (0.63:0.59:0.51). These differences, however, are not statistically significant (F-value of 1.51 on 2 degrees of freedom; p-value of 0.24).

- Respondents from the Energy sector are on average more consistent (0.63) than respondents from the Rail sector (0.6), who in turn are more consistent than respondents from the Telecoms sector (0.57). Respondents from the Competition sector are last, with an average consistency of 0.45. Again, however, these differences are not statistically significant (F-value of 1.86 on 3 degrees of freedom; p-value of 0.16).

4.4.2. Reliability

It is possible for respondents to be consistent, but still be very bad judges. Respondents might, for example, consistently rate regulators in the opposite way to that given by all other respondents. In order to test this, we can calculate measures of inter-respondent *reliability*.

Specifically, we can calculate values of Cohen's kappa statistic (Cohen, 1960) for each pair of respondents, and then calculate the average reliability level across all pairs of respondents. Using a statistic such as Cohen's kappa is preferable to using simpler

measures of agreement, such as the percentage of occasions on which respondents agreed. Raw measures such as this fail to take account of the possibility that respondents might agree by chance, and also that the three response categories are ordered, so that disagreement between one rater who prefers A to B, and another rater who judges them to be equally good, is less than the disagreement between one rater who prefers A to B and another rater who prefers B to A.

Cohen's kappa has a maximum value of one, which indicates perfect agreement. Values of zero indicate that the observed agreement between two raters is no better than chance. Values of less than zero indicate that raters actively disagree with each other. Specifying acceptable levels of kappa is difficult because the values of kappa themselves change based on the number of possible responses open to raters. However, one rule-of-thumb scale is provided by Landis and Koch (1977), and is listed in Table 4.3.

Table 4.3.: **Kappa: rules of thumb**

$\kappa < 0$	No agreement
$0 < \kappa < 0.2$	Slight agreement
$0.21 < \kappa < 0.4$	Fair agreement
$0.41 < \kappa < 0.6$	Moderate agreement
$0.61 < \kappa < 0.8$	Substantial agreement
$0.81 < \kappa < 1$	Near-perfect agreement

For our data, the value of kappa averaged out over all rater pairs is 0.22, indicating fair agreement. However, it would clearly be desirable if we could improve the level of reliability between respondents.

One strategy for doing so is to eliminate respondents progressively based on their contribution to the average value of kappa. That is, for each respondent we calculate the value of kappa which would obtain if that respondent was not included. We then identify the largest improvement in kappa, and remove the associated respondent. By proceeding iteratively, we can identify a subset of respondents who are collectively more reliable.

Here, we adopt a threshold of $\kappa = 0.4$, which is equivalent to requiring moderate agreement of our respondents. By proceeding iteratively, we identify ten respondents who drag down the value of kappa most.

Note that these respondents are not necessarily the least consistent respondents. Some of these respondents are extremely consistent when rating regulators. However, they

drag down the reliability statistic because their consistent judgements are at odds with the rest of the group. This may be because they are, in some sense, poorer judges; or it may be because their judgements are affected by particular structural factors which we identify later.

4.4.3. Strategies for dealing with consistency and reliability

In the two preceding sections, we tested for consistency and reliability. In the first case, we could not reject any respondents for statistically significant levels of inconsistency. In the second case, we were able to identify ten respondents whose responses dragged our reliability statistic below a certain arbitrary level. In neither case are there conclusive reasons for excluding participants. Additionally, when considering reliability, it is possible that levels of disagreement could reflect structural characteristics which can be accounted for.

Our strategy for dealing with consistency and reliability is therefore to perform all of our analyses four times:

- on the full data-set;
- on the full data-set, but weighting all responses by respondents' levels of consistency, as measured by their ζ-statistic;
- on a restricted data-set, excluding the ten respondents most responsible for low levels of κ;
- on a restricted data-set, excluding the ten respondents most responsible for low levels of κ, and weighting the remaining responses by respondents' levels of consistency, as measured by their ζ-statistic;

and then to demonstrate that these different analyses deliver substantially the same conclusions about the relative perceived quality of regulators.

4.5. Modelling perceived quality: baseline model

We can analyse this pattern of pairwise comparisons to establish a general ranking of regulators' perceived quality. This ranking will naturally be very uncertain, given the limited number of respondents, the possibility for respondents to compare regulators on their basis of their own subjective criteria, and the possibility of certain types of re-

spondent bias. In this section, we present first a baseline model of the perceived quality of regulators, before moving on to test for respondent bias.

The model we use to analyze this pattern is called a Bradley-Terry model (Bradley and Terry, 1952). These models estimate the probability of one contestant, i with a given level of quality (λ_i) beating another contestant, j.

$$logit[Pr(i\text{beats}j)] = \lambda_i - \lambda_j$$

Here, our contestants are regulators, and the contest is the pairwise comparison made by the respondent. The greater the difference between the quality of regulator i and the quality of regulator j, the more likely it is that regulator i will 'beat' j, or, equivalently, be chosen by our experts. By maximum likelihood estimation techniques, we can find, for each regulator, the value of λ that makes the observed pattern of 'victories' and 'defeats' (that is, the observed pattern of expert choices) most probable.

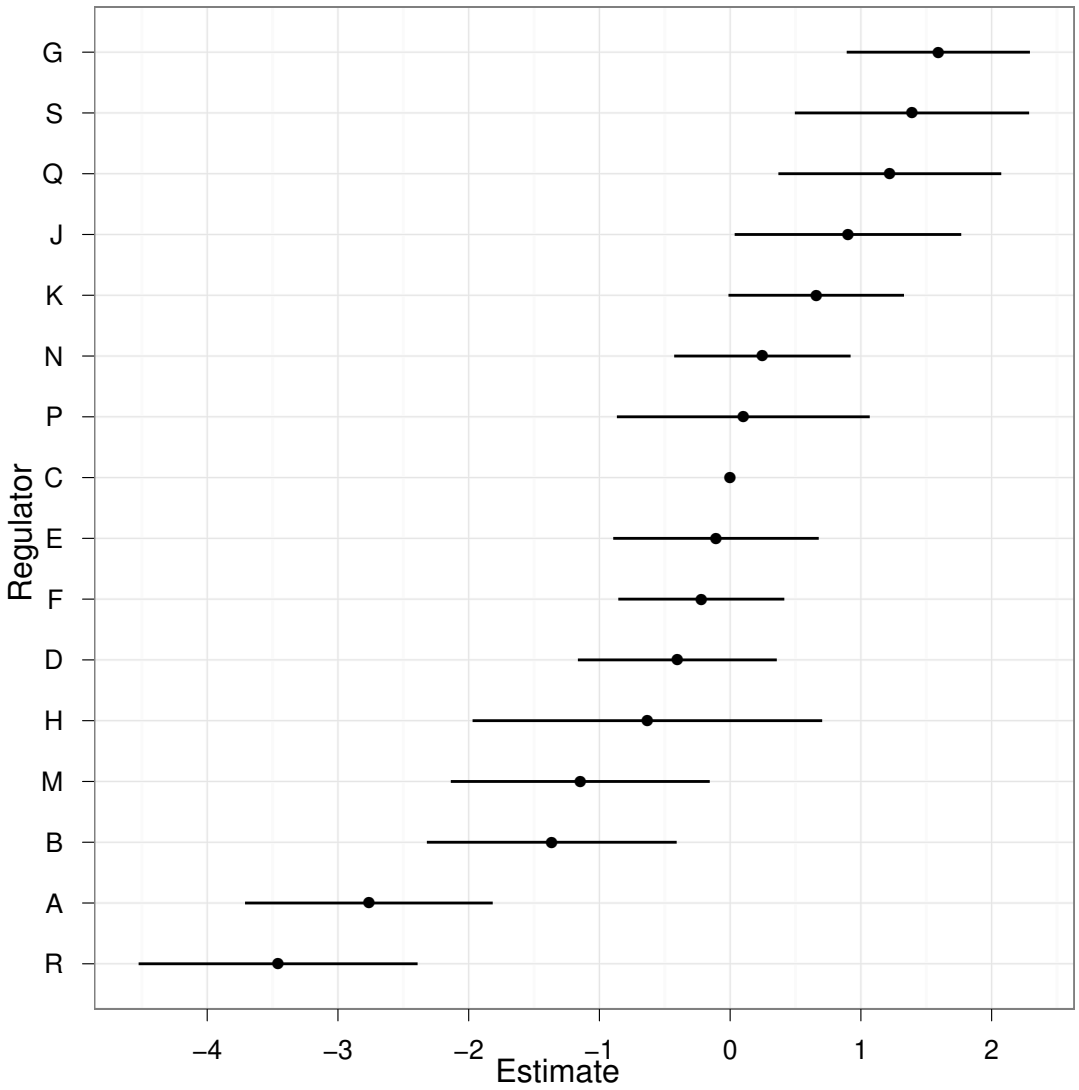
These estimated values are point estimates, but they come with a degree of uncertainty. Consequently, it is important to interpret them with due diligence.

Figure 4.2 shows the value of regulators' perceived quality in this baseline model, using the full data-set. For each regulator, we plot the point estimate of the regulator's quality (given by the dot) as well as the 90% confidence interval. Thus, we can be 90% confident that the true value of the regulator's quality lies within this range. There is one exception: regulator C is used as the reference category, and its value is set to zero. (This has no implications for Regulator C's relative position: were it really perceived to be the best, the values of the other regulators would all be below zero). No confidence interval is given.

We can see that some regulators have broader confidence intervals than others. This may reflect either the fact that respondents were genuinely uncertain about how to judge the quality of these regulators, or that these regulators were involved in a limited number of comparisons (this is the case for the German regulators, for example).

In terms of the separation between the regulators, we can be very confident that there are differences between the top four regulators and the bottom four regulators; the ranking of the middle-placed regulators, although still most likely to be in the order given, is more uncertain given the way in which the confidence intervals overlap.

Figure 4.2.: **Regulator perceived quality, baseline model**



The figures presented in Figure 4.2 show the parameters as estimated on the full data-set, weighting all respondents equally. However, as noted earlier, concerns about respondent reliability and consistency mean that it is important to test whether these results are robust to alternate specifications, in particular alternate specifications which exclude 'unreliable' respondents and place greater weight on consistent respondents. A scatter-plot matrix of these alternate measures is shown in Figure 4.3, which plots the results under the four different specifications discussed above: the full data-set, the data-set weighted by respondent consistency, a restricted data-set excluding ten 'unreliable' respondents, and a restricted data-set weighted by respondent consistency. The scatter-plot matrix shows that these alternate specifications barely affect the order and relative positiong of the sixteen regulators. All of the correlations between these variables are extremely high, at greater than 0.9. The only outlier is Regulator M, which receives a higher rating in the full analysis than it does in the restricted analyses.

Given the high correlations between these different measures, and since measures calculated on the full data-set are more efficient (have narrower confidence intervals) than measures calculated on the restricted data-set, only measures based on the full data-set are plotted here; all substantive interpretations of these measures refer to measures calculated on the full data-set.

4.6. Modelling perceived quality: biases

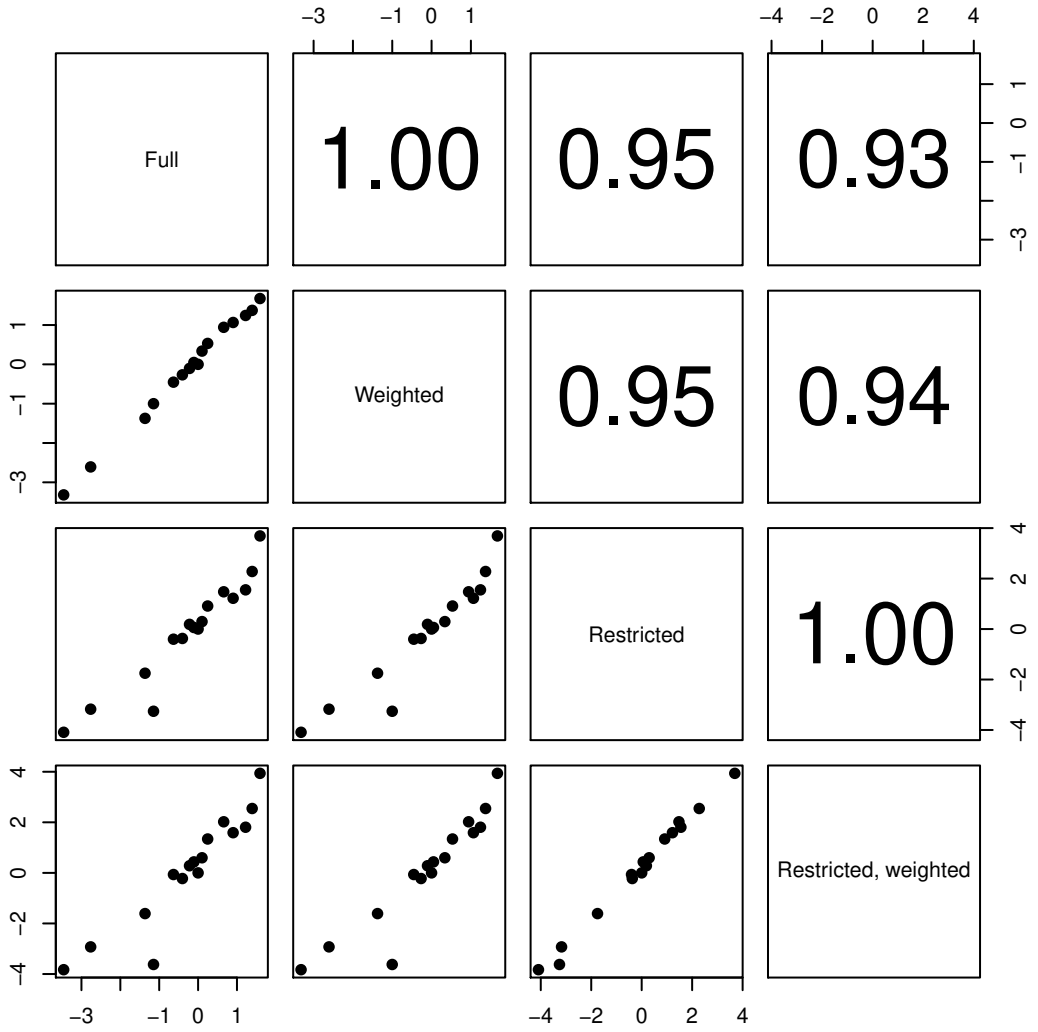
The estimates presented in Figure 4.2 assume that respondents are unbiased (though perhaps noisy) judges of quality. However, it is possible that respondents suffer from a number of biases. In advance planning for this exercise, we identified a number of potential biases which it would be wise to control for:

- 'Regulator-boosting': respondents working for regulators will be more likely to choose their employer over any other organisation, irrespective of the true level of quality;

- 'Regulatee bashing': respondents working for regulated organizations will be more likely to choose any other organisation over 'their' regulator, irrespective of the true level of quality. This may be because regulatees wish to discredit their regulators, or simply (and more likely) because they believe the grass is greener elsewhere;

- 'Nationalism': respondents from country A will be more likely to choose organisations from their country over organisations from any other country, irrespective of

Figure 4.3.: **Regulator perceived quality, alternate specifications**



the true level of quality.

These biases can be incorporated in to the Bradley-Terry model by simply adding an additional parameter which captures this effect. Thus,

$$logit[Pr(i\text{beats}j)] = \lambda_i - \lambda_j + \delta Z$$

where Z is a matrix with columns representing non-player specific advantages, rows representing contests, and cell entries taking a value of 1 if player i has the advantage, and -1 if player j has it. Thus, if we are testing for regulator-boosting, then the entry in the appropriate cell in Z would be '1' if the respondent works for the first-named regulator, '-1' if the respondent works for the second-named regulator, and zero otherwise.

Table 4.4 shows estimates of the size of these biases, along with tests for statistical significance. We discuss each of these in turn.

Table 4.4.: **Sources of bias**

	Bias	Estimate	Std. Error	z value	Pr(>|z|)
1	Nationalism	0.48	0.35	1.36	0.17
2	Boosting	1.52	0.46	3.28	0.00
3	Bashing	0.37	0.37	1.01	0.31

4.6.1. Nationalism

When we test for nationalism, we expected a positive coefficient. The coefficient reported in Table 4.4 is indeed positive, but is not statistically significant at standard levels of significance. This may result from differences in nationalism across respondents. Specifically, respondents from some countries may have very high views of regulatory quality in their own countries, whilst respondents of other countries may have a very pessimistic view. Britons, for example, may be very bullish about regulatory quality in the UK compared to other respondents, whilst Dutch respondents (of which they were fewer in our sample) may be particularly pessimistic about regulatory quality in the Netherlands.[12] Other effects may also be at work. Respondents may simply know more about regulators in their country – which would not necessarily have a strong net positive or negative effect on pairwise comparisons. Because the coefficient is not sta-

[12] This example is only illustrative: we are not suggesting that this is in fact the case.

tistically significant, and because uneven responses like this are eminently possible, we do not correct for this particular bias.

4.6.2. Regulator-boosting

When we test for regulator-boosting, we can see from Table 4.4 that respondents do suffer from bias when answering questions about their own regulator. The effect is positive (that is, respondents are more likely to find in favour of their own regulator) and significant. The effect is also quite large: if the only respondents we listened to were biased in this way, then a regulator might shift from the bottom quartile of Figure 4.2 to the top quartile. Fortunately, responses affected by this bias are a minority of total responses, and there is no reason to suggest that respondents boosting their own regulators are biased when evaluating contests in which they have no stake.

4.6.3. Regulatee bashing

When we test for regulatee bashing, we expected to see a negative coefficient in Table 4.4. That is, we expected regulatees to be less likely to choose their own regulator. In fact, regulatees are more likely to choose their own regulator in national or sectoral comparison. Note though that this effect is not sigificant, and thus will not be taken into consideration at this stage.

4.6.4. Corrected results

Figure 4.4 shows results corrected for the 'boosting' bias. The major loser from this correction is regulator M, which loses 0.28 points. The major beneficiary from this correction is regulator N, which gains 0.41 points. The correlation between the two measures (the corrected estimates and the original, uncorrected estimates) is, however, extremely high, at 0.99.

Just as before, we can test whether the figures shown in Figure 4.4 are robust to alternate specifications. Figure 4.5 again plots a scatterplot matrix of the measures of perceived quality as calculated using different sets of data. Once again, the correlations between different types of measure are extremely high, and greater than 0.9.

Figure 4.4.: **Corrected results**

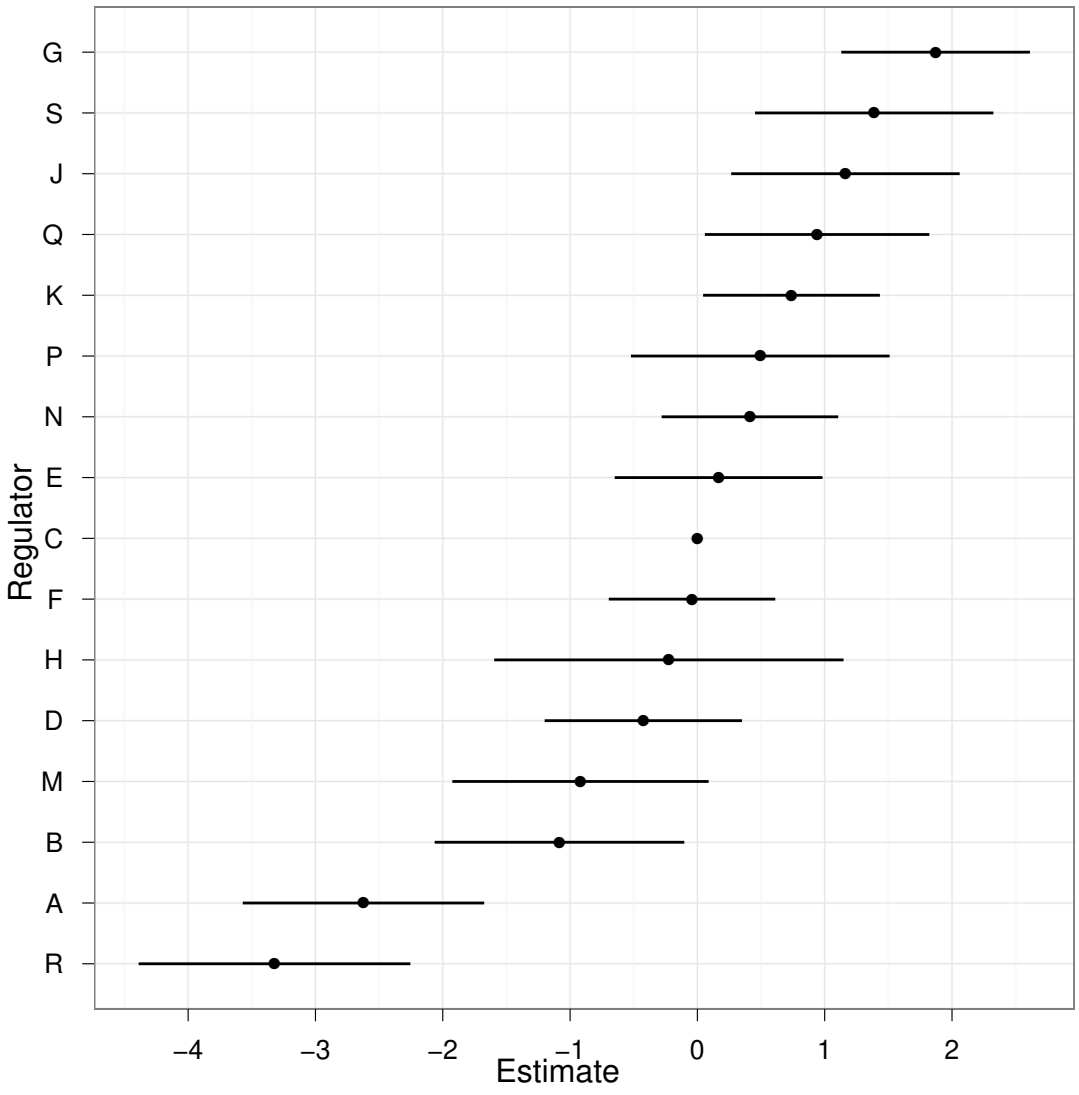
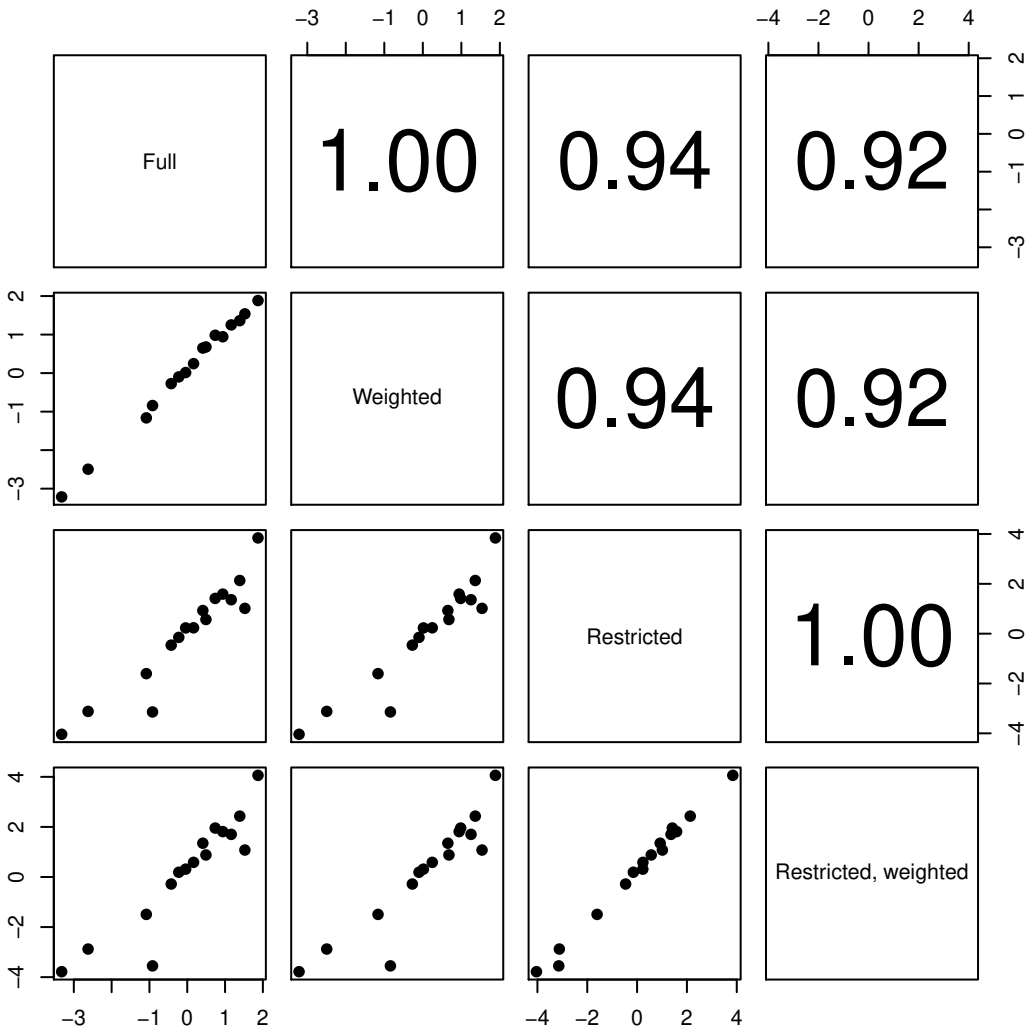


Figure 4.5.: **Regulator perceived quality, alternate specifications**



4.7. Discussion of results

If the results shown in Figure 4.4 are reliable indicators of regulators' perceived quality, what explains why some regulators are perceived as being of better quality than others? Whilst a full analysis of this will follow with our measures of independence and accountability, we can show how perceived quality varies according to country and sector.

Figure 4.6 first shows how regulators' perceived quality varies by country. The UK is considerably ahead of all other countries, followed by the Netherlands and Germany, though in these two cases we can only base our conclusions on the two multisectoral regulators in these countries.

Figure 4.7 then shows how regulators' perceived quality varies by sector. The disparities between sectors are less marked than the disparities between country. More specifically, whilst there is a statistically significant difference between the perceived quality of regulators according to country (F-value of 4.33 on 4 degrees of freedom, p = 0.02), the same cannot be said of quality by sector (F-value of 1.31 on 3 degrees of freedom, p = 0.32). Nevertheless, Telecoms edges out Competition, which in turn is ahead of Energy. Rail is the last placed sector.

These indicative results are a precursor to a fuller analysis of perceived quality according to the independence and accountability of regulators. For the moment, it suffices to restate the main points of this chapter.

We have presented an overview of an exercise designed to elicit the perceived quality of regulators in five countries and four different sectors. We have shown that experts are fairly consistent in their pairwise comparisons of regulators. We have modelled these pairwise comparisons using a Bradley-Terry model, and presented initial results of the perceived quality of regulators. We have then tested for the presence of three possible biases when using such data, and have presented corrected results which account for the regulator-boosting seen in the data.

Figure 4.6.: **Regulator perceived quality by country**

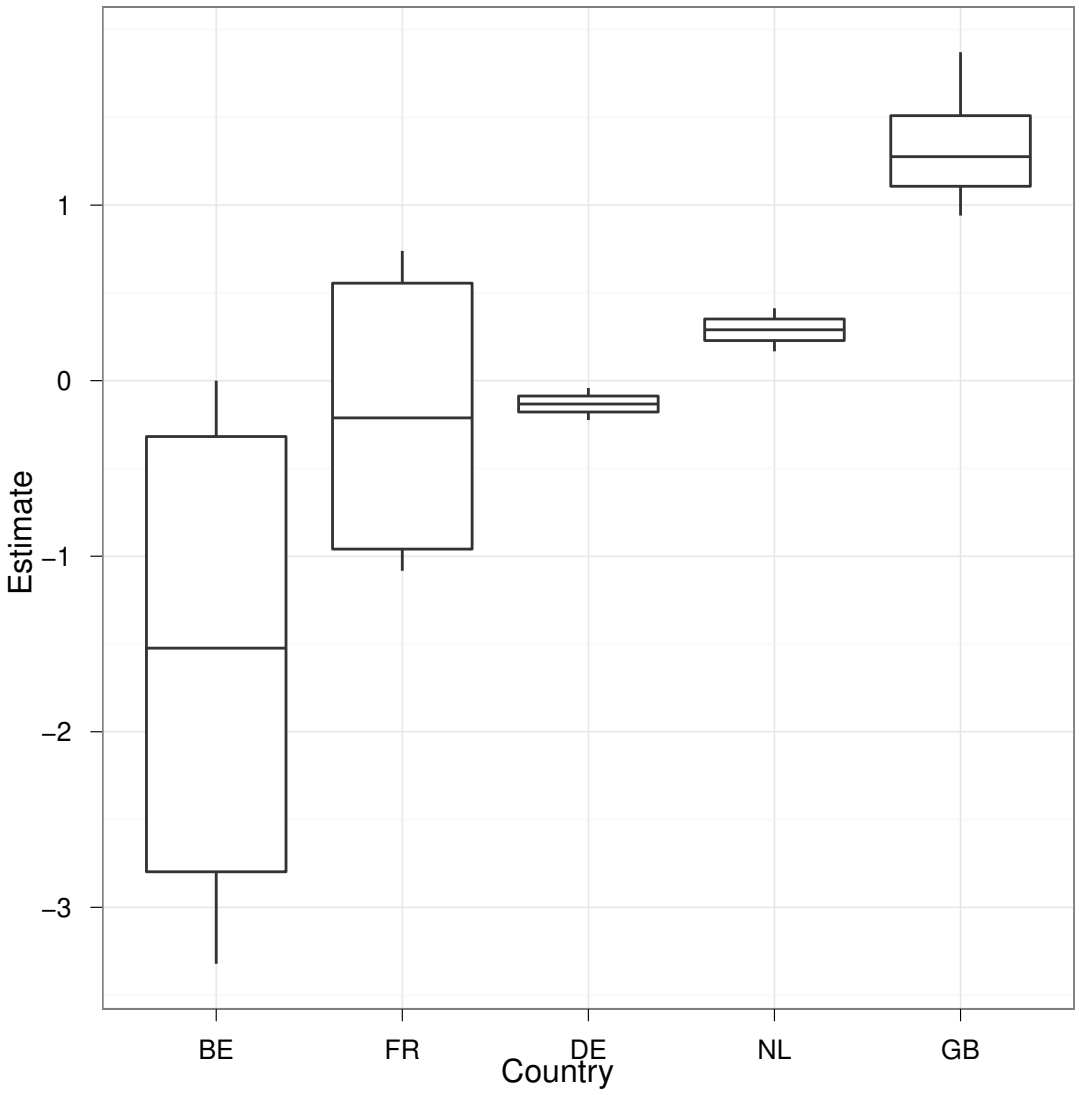
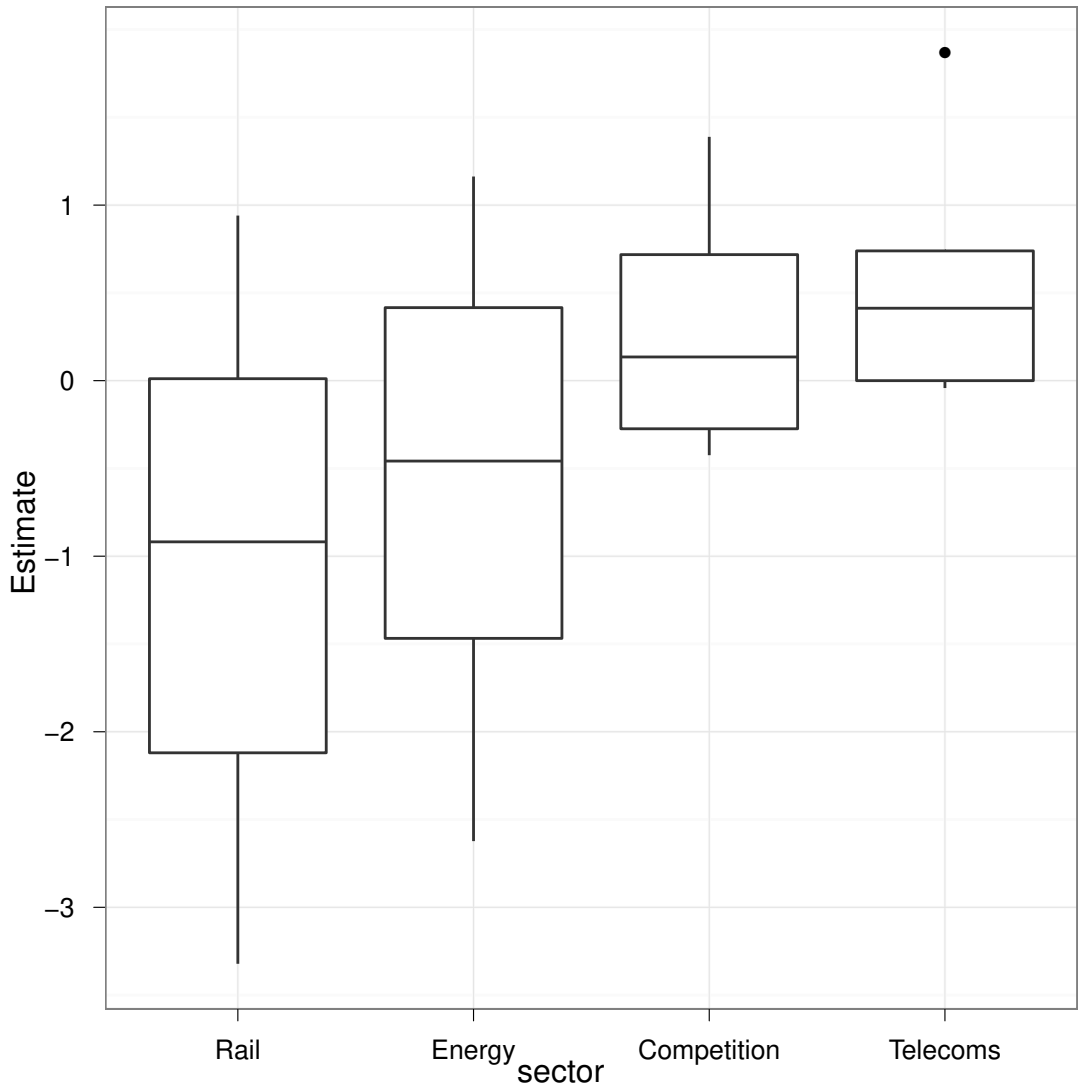


Figure 4.7.: **Regulator perceived quality by sector**



5. Empirical links between independence, accountability and perceived quality

In this section, we discuss the links between independence, accountability and perceived quality – though our main focus is on explaining levels of perceived quality. We begin by discussing a potential confounder, namely the resources available to each regulator. We discuss how to measure this, before moving on to show the bivariate relationships between each of the three main features we are interested in, plus levels of resources. We then present a series of linear regression models which attempt to take account of multiple variables at a time. We close by discussing the implications of these models.

5.1. Regulator resources

Earlier, we discussed differences in the perceived quality of regulators, as revealed by pairwise evaluations made by regulators, regulatees, and academics. One potential explanation of these differences lies in the resources available to each regulator. Some of the regulators we consider here have very large budgets in absolute terms, whereas some have rather small budgets. It is possible that these differences are driving differences in perceived quality, either because funding really does matter for quality, or because those comparing regulators opt for the larger, better-funded, or more visible, regulator. Whilst the link between funding and quality of work seems intuitive – more money can pay for more comprehensive and better economic and legal analysis – we cannot rule out the second possibility.

We will therefore examine the connection between the resources available to each regulator, and its perceived quality. Note that we examine the resources available to each regulator in an absolute sense. That is, we make no attempt to evaluate whether the resources available to each regulator are in any way 'adequate' to the tasks and competences it has been assigned. Multisectoral regulators are likely to have larger budgets than single-sector regulators, but the increase in funding may not be proportionate to the increase in competences.

To measure and make comparable the resources available to each regulator, we collected information on the size of the budget, expressed in millions of euros, for each regulator, and on the number of full-time equivalent staff for each regulator. These fig-

ures were generally taken from the annual reports of each regulator, and typically refer to the financial year 2010/11.

We chose to combine these two sources of information to produce an overall measure of the resources available to each regulator. We did this for two reasons. First, it is possible that some regulators may choose to operate in either a capital-intensive or labour-intensive fashion. Taking just information on the budget of the regulator, or on the number of the staff, might then give a misleading impression of the available resources. Second, because the number of regulators we are dealing with is limited, there are technical difficulties involved in including both of these sources of information separately. Visually, we would not be able to plot simple bivariate relationships between resources and perceived quality; statistically, including both of these (correlated) variables in a regression with a limited number of cases would mean that we were unlikely to obtain unbiased and/or efficient estimates.

Consequently, we produce a single composite measure of resources as follows:

- We took the *log* of budget (in millions of euros) and staff (in full-time equivalents). We took the log because both staff and budget show decreasing marginal utility, and thus increases from a few staff to a moderate number of staff matter more than increases from a moderate number of staff to a large number of staff;

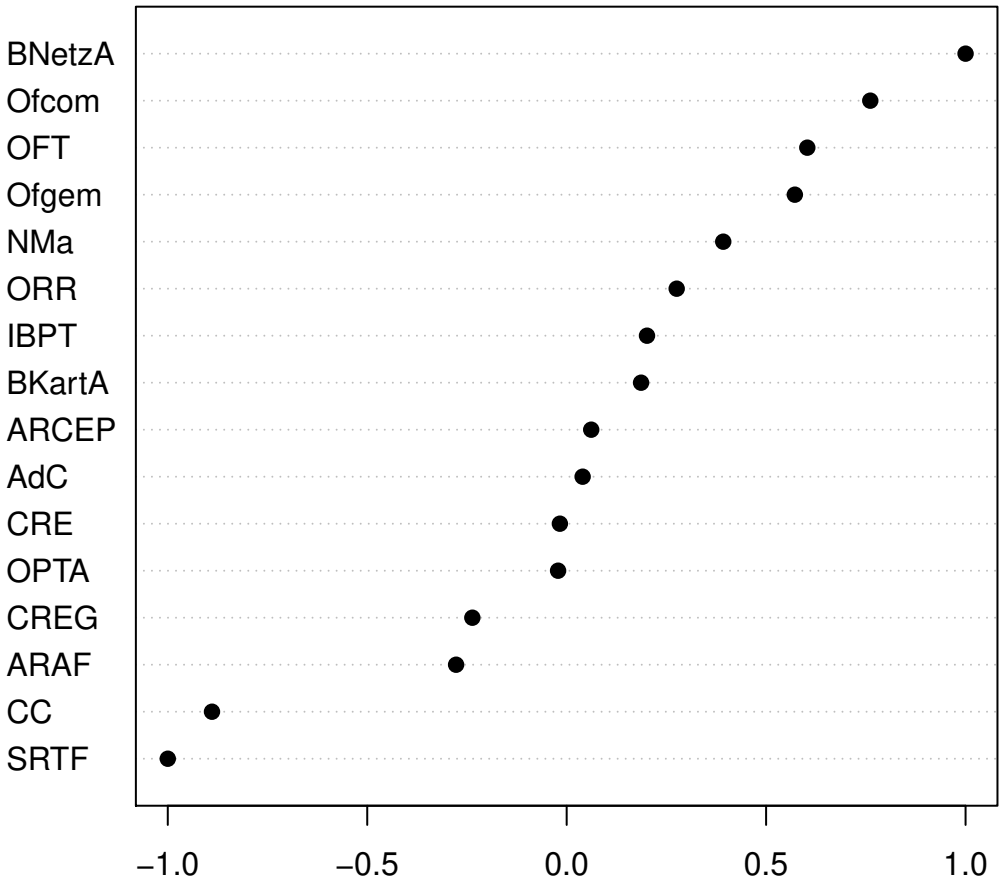
- We *scaled* each variable to have the same mean (zero) and the same standard deviation (one). We did this in order to combine these two variables which are measured on different scales (euros and persons);

- We took the *average* of the two figures, and *transformed* the data so that the smallest value was equal to minus one, and the largest value was equal to positive one.

Although we motivated combining these two sources of information in part because we were worried about capital-intensive or labour-intensive outliers, the Spearman rank correlation between the two variables of 0.98 is very strong and positive.

Information on this summary measure of resources is plotted in Figure 5.1. It shows clearly a clustering of seven or eight regulators around the zero point (the mean of our standardized variable), with a number of outliers to either side.

Figure 5.1.: **Regulators listed by resources**



5.2. Resources, independence and accountability

It is useful to investigate the relationship between all three of the 'input' variables we are using to explain quality. If all three of these inputs are positively related, then neither regulators nor policy-makers face any trade-offs in pushing for legislative changes which might affect resources, independence, or accountability. If, by contrast, some of these variables are negatively related to each other (as some people have assumed in the case of independence and accountability), then those decision-makers do indeed face a trade-off.

Figure 5.2 therefore plots the relationship between these different inputs. The Pearson correlation of 0.46 between independence and accountability is strong and positive; a similar figure of 0.46 is obtained for the relationship between independence and resources. The strongest correlation (0.82) is between the degree of accountability and the quantity of resources available to the regulator. This correlation might suggest that politicians are happier to disburse money when they believe that they can hold recipients accountable. However, there would be no reason to expect such a relationship with respect to resources derived from own revenue sources, such as industry levies.

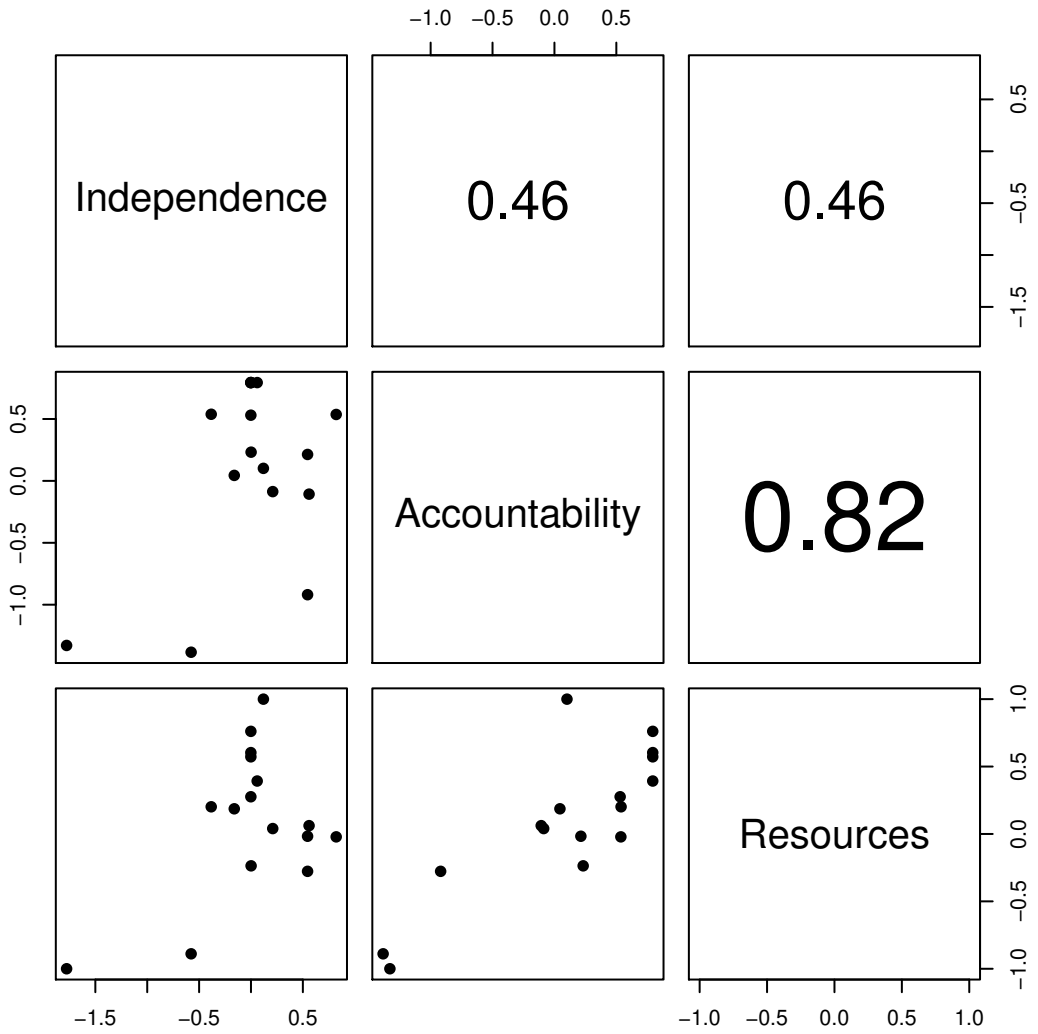
5.3. Quality and inputs

We now turn to discussing the bivariate link between perceived quality and our inputs. Considering first the link between quality and resources, the Pearson correlation between the two variables This Pearson correlation of 0.71 is very strong and positive.

If resources were the exclusive non-idiosyncratic determinant of quality, and if quality was judged correctly by those who compared regulators, then we would have 8 regulators which are delivering better-than-expected performance, and 8 regulators which are delivering poorer-than-expected performance. The three most over-performing regulators are D (1.49), K (0.91), and G (0.75). Of course, all of these judgements depend on the assumptions that resources are the exclusive non-idiosyncratic determinant of quality, an assumption which we presume, in virtue of our emphasis on independence, to be false. It is to the links between independence and accountability, on the one hand, and perceived quality, on the other hand, to which we know turn.

The Pearson correlation between these two variables of 0.48 is strong and positive. If independence was the exclusive non-idiosyncratic determinant of quality, and if quality was judged correctly by those who compared regulators, then we would have 10 reg-

Figure 5.2.: **Relationship between different inputs**



ulators which are delivering better-than-expected performance, and 6 regulators which are delivering poorer-than-expected performance. The three most over-performing regulators are G (1.96), S (1.48), and J (1.25).

Finally, we turn to the bivariate link between quality and accountability. The Pearson correlation of 0.62 is strong and positive.

5.4. Overall results

Thus far, we have looked at bivariate relationships between quality, on the one hand, and our three 'inputs' on the other hand – resources, independence, and accountability. We have seen that all three of these inputs are positively related to quality, and indeed they are all positively related to each other. We can now move on to estimating the effects of these different inputs when they are considered at the same time.

It is natural to model the impact of these inputs on quality through linear regression. Accordingly Table 5.1 lists three regression models of quality based on different combinations of the three inputs.

Unfortunately, because of the limited number of regulators in our sample (16), it is both difficult to achieve results that are statistically significant, and undesirable to include all three predictors. Most rules of thumb usually suggest between ten and twelve cases per predictor, a rule of thumb which would permit only bivariate tests of the relationship between inputs and quality. Nevertheless, the models shown suggest that, across the three different pairwise combinations of these three inputs, there is a significant relationship between at least one variable in each model.

Resources have a significant positive impact on quality in the first model, alongside independence. Resources also have a significant positive impact when considered alongside accountability, although here the relationship is only significant at the 10% level. Finally, accountability has a significant positive impact when considered alongside independence.

None of these models is likely to be the 'correct' model of quality, but the three models taken together do suggest that there are positive relationships between these three inputs and quality which are not epiphenomenal. In terms of the 'best' model, the first model, relating quality to the resources available to the regulator, and the degree of independence enjoyed by the regulator, is the most successful at explaining variation

Table 5.1.: **Regression models of quality**

	(1)	(2)	(3)
(Intercept)	−0.256	−0.277	−0.184
	(0.258)	(0.264)	(0.282)
Resources	1.608*	1.594†	
	(0.543)	(0.865)	
Independence	0.455		0.573
	(0.491)		(0.545)
Accountability		0.221	0.972*
		(0.642)	(0.448)
R-squared	0.541	0.515	0.436
adj. R-squared	0.470	0.440	0.350
sigma	1.006	1.034	1.115
F	7.660	6.904	5.031
p	0.006	0.009	0.024
Log-likelihood	−21.144	−21.583	−22.787
Deviance	13.167	13.910	16.169
AIC	50.289	51.167	53.574
BIC	53.379	54.257	56.664
N	16	16	16

Note: * = results statistically significant at the 0.05 level of significance; † = results significant at the 0.1 level.

in quality, able to explain nearly half of the variance in quality, as shown by the adjusted R^2 figure given in the table.

5.5. Conclusions

We began this section by discussing in detail one factor which might potentially confound relationships between independence, accountability, and quality, namely the level of resources available to the regulator. We indicated how we measured this, and present summary information on the resources available to each regulator in our sample. We then went on to present simple bivariate relationships between the four variables we are interested in: resources, independence, accountability, and perceived quality. We showed that all four of these items were positively related, and that the relationships were statistically significant. In closing, we presented a number of regression models which attempted to estimate the impact of these variables upon perceived quality, considered jointly. Although not all of the estimates presented in these models were significant, this was due to a limited sample size; generally, the results strongly suggest (but do not confirm) an important relationship between independence and accountability, on the one hand, and perceived quality on the other hand, even when these two variables are considered together.

6. Conclusions and policy recommendations

In this section, we summarize our findings. We follow the structure of the report, and discuss independence, accountability, and perceived quality, before going on to summarize our findings regarding the links between these three properties in greater detail.

In terms of **independence**, we discussed two main aspects of independence – independence from politics and independence from market players – and how these aspects could be considered either *de jure* or *de facto*. We then went on to build an index of independence in terms of several different items. Some of these items could be answered in a simple yes or no fashion. 'Is the regulator described in law as politically independent' would be one example. Some of these items were more complex, and had multiple possible responses which could be ordered in terms of increasing independence. The length of term of the heads or chief executives of regulatory bodies was one such item, with longer terms corresponding to more independence. Using a particular statistical model, we mapped these items in terms of their contribution to increasing independence. We gave two numbers for each item: its location, on a scale from low independence to high independence, and its discrimination: its ability to separate regulators with otherwise similar levels of regulators. We found that some items had relatively low locations, such as clear statements of the regulator's independence, and a bar on ministerial instructions. We found that other items were much more demanding, such as provisions on making renewal of appointments more difficult. We then used this index to score the independence of the regulators in our sample.

In terms of **accountability**, we noted that accountability is a relationship between the regulator and other actors, usually including national politicians and members of the executive but also potentially including market participants and the public, the judicial system, and the European level. We then analyzed accountability in a very similar way to the way in which we analyzed independence. We constructed an index of accountability using several different items. We excluded some items from this index because there was no variation between regulators, which made it impossible to use them in our index. With the items that were left, we once again 'mapped' the location and discriminating ability of each of these items. We were left with an index that included certain procedural guarantees of accountability as the bedrock of accountability, followed by provisions regarding the publication of information, provisions regarding discovery pro-

cedures by politicians and market players, and, finally, provisions regarding the evaluation of the regulator's work. Once again, we used this index to score the accountability of the regulators in our sample. In this case, we found that it was much more difficult to rank regulators in terms of accountability. This greater uncertainty is in part a result of the model, and in part a result of the regulators we examined. The four UK regulators, for example, had very similary provisions relating to accountability, which makes it extremely difficult to differentiate between them, and in turn reduces the variability of the patterns we see in the data, patterns we exploited to construct our index.

It is interesting to put side by side our findings regarding independence and accountability, according to the entity that relates to the regulatory agency.

First of all, as regards market players and other stakeholders, and more broadly the general public, we see that independence is firmly established since the introduction of the current regulatory schemes across Europe, while at the same time a number of accountability measures benefit these actors. Most measures at the four stages identified in our report (procedural, information, discovery and evaluation) benefit these actors.

Secondly, as regards the national legislative and executive powers, we can observe the progressive emergence of a principle of independence of regulatory agencies in network industries, which has now taken hold in the latest revisions of the electronic communications and energy directives, in step with recent rulings by European courts. This slower evolution (compared with the independence from market players) can be explained in part by difficulties arising from the need to fit independent agencies within the constitutional and administrative law frameworks of EU Member States. Here theory is taking time to catch up with practice, where regulatory agencies have enjoyed independence for some time already. In parallel to this longstanding debate, a number of accountability measures have been put in place, to provide a counterbalance to independence. Here as well, many of the measures identified in our report – in particular the production of periodical activity reports, the appearance before parliamentary committees, the conduct of performance evaluations – create accountability towards the legislative and executive powers. As we noted, the accountability measures which would most squarely conflict with the independence of the agency – prior ministerial approval, the issuance of ministerial guidelines, the dismissal of the head of the agency by the ministry – are not necessary to create accountability. In other words, a high level of accountability does not need to come at the expense of independence. In the end, accountability is ensured without the need for hard sanctions from the legislative

or executive power.

Thirdly, as regards the national judiciary power (national courts), we note that a number of accountability measures, especially at the procedural stage, benefit the judiciary. In contrast with the legislative or executive, however, the judiciary holds the power to sanction the regulatory agency through judicial review, and that power has been, and continues to be, used across the Member States and sectors surveyed in this report. At the same time, neither European nor Member State law provides for independence of the regulatory agency from the judiciary. At first sight, the relationship with the judiciary – with respect to both accountability and independence – seems to be at variance with what was outlined in the previous paragraphs regarding the legislative and executive or the market players and general public. Yet, without prejudice to further research on this point, we think that it is not. The position of the judiciary is simply different. Whereas the legislative and executive powers, as well as market players and the general public, can be considered as stakeholders or principals in the regulatory process, with the regulatory agency being their agent, the judiciary does not act as a stakeholder or principal. At least as a matter of constitutional theory, the judiciary is not pursuing its own objectives and its own agenda in regulatory processes, but rather enforcing the law.

The judiciary would then better be considered as another agent, endowed with its own independence and active besides the regulatory agency. The independence of the judiciary can even be seen as a further safeguard for the independence of the regulatory agency from other actors. The judiciary is entrusted with its own mission – ensuring the compliance with the law – which can lead to conflict with the regulatory agency, in which case judicial review provides for a rule of conflict resolution where the judiciary has the last word.

In this sense, judicial review would constitute a means for the other actors – in particular the market players and the general public – to sanction the regulatory authority, albeit through the mediation of the judiciary. Furthermore, given that the regulatory authority is independent from these other actors, it might explain that judicial review is conducted according to a standard of marginal review, which would reflect the independence of the regulatory authority towards those principals whose interests are protected by judicial review.

Fourthly, as regards peers, i.e. the other regulatory agencies, we remark the presence of accountability mechanisms, primarily via the work of regulatory networks such as BEREC or ACER, without any measure of independence. Here we can consider that

since all regulatory agencies neither instruct nor sanction one another, they do not need to be independent from one another. In addition, to put it metaphorically, they sit on the same side of the fence.

Finally, as regards the European Commission, we note a number of accountability measures, including the ability for the Commission to intervene against the regulatory agency, be it via sector-specific means (for instance, the procedures of Articles 7 and 7a of Directive 2002/21) or via infringement proceedings under the TFEU. At the same time, there are no measures to ensure the independence of Member State regulatory agencies towards the European Commission. This raises some interesting issues for further research. Indeed the position of the European Commission cannot be assimilated to that of the national courts. The Commission is a principal in its own right, with its own policy agenda (or that which it derives from interplay with the other EU institutions), which it seeks to accomplish through national regulatory agencies. Similarly, the Commission cannot easily be reduced to just one node in a regulatory network with the national agencies, considering that it holds specific means of intervention that place it above the agencies rather than amongst them. As is well known, the issue of independence of regulatory agencies towards the Commission, under EU law, is left to the outdated Meroni doctrine, which rests on the formalistic policy/implementation distinction.

In terms of **perceived quality**, we discussed the results of a peer review exercise conducted late last year in which participants were asked to make pairwise comparisons between regulators operating in their country, and regulators operating in their sector. We checked the consistency of these comparisons, both overall and between the different types of respondent (academics, regulators, and regulatees), and found that levels of overall consistency were acceptable. We then went on to test for a number of biases in responses (such as the tendency to 'boost' one's own regulator, for regulators, and the tendency to 'bash' one's own regulator, for regulatees), and corrected for these. We then used these pairwise comparisons to construct a third index, this time an index of perceived quality.

In terms of the **links between independence, accountability, and perceived quality**, we showed that there is both (1) a statistically significant and positive link between independence and perceived quality, and (2) a statistically significant and positive link between accountability and perceived quality. What is more, independence and accountability are themselves positively related. This suggests that robust independence and accountability mechanisms can co-exist and contribute to the effectiveness of a

regulator. Evaluating the joint impact of independence and accountability is difficult because of the limited number of regulators in our sample, and because of the important confounding effect of the level of resources available to the regulator.

Our research provides empirical support for the basic approach chosen in the EU for the design of regulatory agencies, namely a combination of independence from the national executive and legislative powers, on the one hand, and accountability to those powers, as well as to European institutions, peers, private actors and courts, on the other hand. A sustainable and fruitful trade-off between independence and accountability is possible.

As far as the policy recommendations stemming from this research are concerned, we make five broad recommendations. One of these recommendations applies to both regulators and policy-makers; three apply principally to regulators; and the last applies to regulatees.

- Our first broad recommendation is that regulators should ask to be granted greater independence and offer more accountability. In making this recommendation, we recognize that we are, in a certain sense, pushing at an open door: regulators are unlikely to advocate for greater and more demanding restraints on their actions. Yet we believe that this recommendation now has added force in the light of our findings regarding the compatibility of independence and accountability, and the link between independence and perceived quality. Our findings help defuse certain arguments against granting greater independence. A common argument found in the political science literature has been that independence and accountability are inimical. That is, there is a trade-off between independence and accountability, such that greater independence comes at the cost of lower accountability, and vice versa. Our research shows that this is not the case, and that indeed independence and accountability go hand in hand. Armed with this finding, regulators can now make the case for greater independence without conceding that this necessarily damages accountability – or, what is perhaps more useful, can now freely accept requests for greater accountability without worrying that this will damage independence. Indeed, our research lends support to the EU regulatory model whereby both independence and accountability are emphasized. Accordingly, public choice literature which fails to factor in accountability (for example, the work of Willam Niskanen) can be dismissed. Similarly, our research shows that legal literature that emphasizes hard mechanisms (sanctions, annulment, etc.) as against softer accountability criteria such as transparency or

disclosure, might be misguided.

This is an important overall judgement, which differs from more stylised facts concerning the impact of independence on sector-specific indicators of quality work such as interconnection rates or reductions in consumer prices. It should be particularly valuable for sectors such as rail where sector-specific metrics of quality are harder to come by. Perhaps more importantly, this finding is easily communicable to policy-makers, since it does not require them to understand the details of any given sector, but requires them only to understand that these are judgements of quality provided by informed experts in regulation in selected sectors.

- Our second broad recommendation is that regulators should pursue increases in independence and accountability as part of balanced packages of changes to their governance. From an analysis of the correlations reported in section 5, it might seem that, since the link between quality and accountability is stronger than the link between quality and independence, regulators should devote all their effort to becoming more accountable, rather than more independent, since more accountability will deliver greater increases in perceived quality. This would be the wrong conclusion to draw, and does not take account of the relationship between independence and accountability. The most natural interpretation of the positive link between independence and accountability is that politicians grant independence if they are satisfied that mechanisms of accountability exist. This link also makes sense from the regulator's perspective. Regulators realize that the principles of accountability and independence are complements that may on occasion pull in opposite directions and therefore need to be balanced to maximize the regulator's effectiveness. Regulators certainly value their independence and non-interference from political institutions, but independence is a double-edged sword: too much can lead to isolation from political processes and udnermine the regulator's ability to influence political decisions, and/or prevent decisions that might undermine the regulator's mission. To overcome the potential downsides of independence and to influence decision making processes, regulators will adopt actions that can be characterised as voluntary accountability measures, including outreach to politicians, improved media relations, and pro-active provision of information to relevant stakeholders.

  Accountability is thus correlated with quality because independence, accountability and quality form a virtuous triangle, which is the basis of the European model of regulatory agencies. Accountability may have an additional effect on quality,

because the act of having to explain actions to diverse publics acts as a spur to regulators, but relying on this effect alone may not be prudent. Pursuing independence and accountability together allows regulators to offer balanced packages of governance changes to policy-makers. Pursuing accountability alone may foreclose future improvements in independence.

We make a minor recommendation which is related to this point of pursuing balanced packages. Regulators should be extremely reluctant to trade increases in the resources available to them for improvements in the governance structures that bind them. In our research we have found it extremely important to control for the different levels of resources available to the regulators. Improved governance structures may therefore not compensate for a freeze in resources.

- Our third broad recommendation is that regulators, in arguing for greater independence and accountability, should focus on 'low-hanging fruit'. By this we mean that regulators who wish to increase their levels of independence should start from the items as mapped in sections 2 and 3, identify which pro-independence (or pro-accountability) provisions they lack, and lobby for the 'missing items' in the order in which they appear in our map. This strategy of focusing on low-hanging fruit makes sense when we relate the order of items given in section 2 not just to their contribution to independence, but also to the willingness of politicians to concede such provisions when drafting the legislation governing the regulator. Thus, focusing on 'missing' low-independence items (such as an explicit statement of independence, or a ban on ministerial instructions) will not only fill in any gaps in the regulator's governance, but will also be easier for ministers or legislators to concede.

- Our fourth recommendation is to regulators who have scored well on accountability and independence. Every regulator in our study can improve on the measures of independence and accountability mapped in sections 2 and 3; but for those regulators which score highly on these measures we recommend a concentration on high-level aspects of accountability which we were not fully able to assess in our index, namely participation in horizontal networks and relationships with the legal system. We were unable to include membership in horizontal networks as an indicator of accountability because all of the regulators in our sample were members of relevant networks. The proper conclusion to draw from this is not that membership in such networks is unimportant for accountability, but rather that explaining and listening to other similarly-situated regulators is widely seen as an essential

component of being a modern European regulator. What may matter more, thus, is not the brute fact of participation in such networks, but the degree to which regulators participate in such networks (possibly overlapping networks), and learn from them. A similar story can be told with respect to accountability to the judicial system: what matters is not the brute fact that regulators are accountable in courts of law, but rather how well they acquit themselves when they find their decisions challenged in a court. These are important aspects of accountability which we have not fully been able to capture in our measurements. If, however, high-independence, high-accountability regulators do seek to cement their reputations as such, improvements in these two aspects would be natural options to pursue.

- Our fifth and final recommendation is a recommendation that regulatees seek to make common cause with regulators in pushing for greater accountability. Certain provisions relating to accountability, such procedural rules and codes of conduct, are of tremendous interest to regulated actors, both *per se* and for their contribution to making decisions evidence-based and thus partly predictable. Greater steps towards accountability thus benefit regulatees as well as regulators.

A. Appendices

A.1. Measures of consistency and reliability for paired comparison data

Where d is the number of circular triads observed, and where n is the number of objects between which a respondent has been asked to choose, then define the coefficient of consistent ζ as follows

$$\zeta = 1 - \frac{24d}{n^3 - n}$$

for odd values of n, and

$$\zeta = 1 - \frac{24d}{n^3 - 4n}$$

for even values of n.

We must, however, account for two particular features of the data:

- Respondents were allowed to express indifference between two regulators. Most calculations of circular triads rely on strict preference relationships only. We therefore correct for ties by following the method outlined in de Vries (1995).
- Respondents were asked to compare two different sets of data – regulators in their sector, and regulators in their country. We must calculate the number of circular triads separately for each set of data. Accordingly, the statistics we present here are averages of the statistics for each set of data.

Table A.1 shows the consistency of each of the respondents. It provides the number of circular triads, the maximum possible number of circular triads, and the value of ζ. Fractional numbers of circular triads result from the adjustment for indifference relations between regulators.

Table A.1.: **Respondent consistency**

ID	Sector	Area	Country	Circular triads	Max possible	Zeta
2	Energy	Regulatee	BE	4.00	7.00	0.43
3	Competition	Regulatee	BE	3.50	7.00	0.50
4	Energy	Academic	BE	4.25	7.00	0.39
5	Rail	Regulator	GB	2.00	7.00	0.71
7	Rail	Regulator	GB	3.25	7.00	0.54
8	Rail	Academic	GB	3.25	7.00	0.54
9	Rail	Regulatee	FR	1.75	7.00	0.75
10	Telecoms	Academic	BE	2.50	7.00	0.64
11	Telecoms	Regulatee	BE	4.75	7.00	0.32
12	Energy	Academic	BE	0.25	7.00	0.96
13	Rail	Regulator	GB	3.50	7.00	0.50
14	Competition	Regulatee	BE	3.75	7.00	0.46
15	Competition	Regulator	GB	3.50	7.00	0.50
16	Competition	Regulatee	FR	3.50	7.00	0.50
17	Rail	Regulator	DE	2.00	5.00	0.60
18	Rail	Regulator	GB	3.75	7.00	0.46
19	Energy	Regulatee	BE	1.50	7.00	0.79
20	Competition	Regulator	GB	4.25	7.00	0.39
21	Telecoms	Regulatee	BE	3.25	7.00	0.54
22	Telecoms	Regulatee	BE	3.00	7.00	0.57
23	Rail	Regulatee	FR	3.25	7.00	0.54
24	Energy	Regulatee	BE	2.00	7.00	0.71
25	Energy	Regulator	NL	1.75	5.00	0.65
26	Rail	Regulatee	GB	0.50	7.00	0.93
27	Telecoms	Regulatee	BE	2.00	7.00	0.71
28	Rail	Regulatee	FR	4.00	7.00	0.43
30	Telecoms	Regulator	BE	4.25	7.00	0.39
31	Telecoms	Regulator	FR	3.25	7.00	0.54
32	Telecoms	Regulatee	BE	2.75	7.00	0.61
33	Telecoms	Academic	GB	1.75	7.00	0.75
34	Energy	Academic	NL	2.50	5.00	0.50
35	Rail	Regulatee	FR	2.75	7.00	0.61
36	Competition	Regulator	BE	4.50	7.00	0.36
37	Telecoms	Academic	NL	1.75	5.00	0.65

A.2. Statistical model of *de jure* independence

Formally, suppose we have i regulators who respond to $j = 1...J$ items, and each item has a number of response categories C_j. Each regulator has an amount of the latent trait θ_i; each item has a difficulty (α_j) and a discrimination parameter (β_j). The *latent response* of each regulator to each item (x^*_{ij}) is therefore

$$x^*_{ij} = \alpha_j + \beta_j \theta_i + \epsilon_{ij}; i = 1, ..., N, j = 1, ..., J$$

which then becomes a manifest response as it is discretized by a number of thresholds for each item, γ_{jc}, for $j = 1...J$ and $c = 0...C_j$, with $\gamma_{j0} \equiv -\infty$, $\gamma_{jC_j} \equiv \infty$, and $\gamma_{j1} \equiv 0$ for the purposes of identification:

$$x_{ij} = c \text{ if } x^*_{ij} \in (\gamma_{j(c-1)}, \gamma_{jc}], i = 1, ..., N, j = 1, ..., J, c = 1, ..., C_j$$

Thus, the more statutory independence (θ) the regulator has, the larger x^*_{ij}, and the more likely it is to clear the thresholds γ_{jc} and answer in a higher response category.

A.3. Coding of legislation

Table A.2.: **accountability items**

Regulator	IBPT	CREG	SRTF	CC	ARCEP	CRE	ARAT	AdC	BNetzA	BKartA	OPTA	NMa	Ofcom	Ofgem	ORR	OFT
Objectives of regulatory policies defined, with priorities	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
Submit annual plan and annual budget	y	y	n	n	n	y	n	n	y	y	y	y	y	y	y	y
Minister can disapprove annual plan, budget	n	n	n	y	n	n	n	n	n	n	y	y	n	n	n	n
Minister can issue general guidelines	n	y	y	n	n	n	y	n	n	y	y	y	n	n	n	n
Provide information to minister, parliament voluntarily, at own discretion	y	y	n	n	y	y	y	y	y	y	y	y	y	y	y	y
Provide information to minister, parliament upon request	y	y	n	n	y	y	y	y	y	n	y	y	y	y	y	y
Present annual reports and financial statements	y	y	n	y	y	y	n	y	y	y	y	y	y	y	y	y
Appear before parliamentary committee, including obligation to answer questions on activities	y	y	n	n	y	y	y	y	n	n	n	n	y	y	y	y
Periodic performance evaluation	y	n	n	n	n	n	n	n	n	n	y	y	y	y	y	y
Minister can dismiss head, board, for performance related reasons	n	n	y	n	n	n	n	n	n	y	y	y	n	n	n	n
Strategies and policy goals explained to stakeholders	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y
Rules on procedures and participation in decisions and rule making	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y	y
Reasoned decisions published	y	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y
Decision making transparent	y	y	y	y	y	y	n	y	y	y	y	y	y	y	y	y
Data underlying decisions published	y	y	n	n	y	y	n	y	y	y	y	y	y	y	y	y
Annual Reports publicly available	y	y	n	n	y	y	n	y	y	y	y	y	y	y	y	y
Code of conduct/guidelines	n	n	n	n	n	n	n	y	y	y	n	y	y	y	y	y
Press releases/newsletters	y	y	y	n	y	y	n	y	y	y	y	y	y	y	y	y
Advisory body with stake holder participation	y	y	n	n	y	n	n	n	n	n	n	n	n	n	n	n
Public consultations with feedback opportunities	y	y	n	n	y	y	n	y	y	y	y	y	y	y	y	y
Judicial appeals with full review	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
Reports to European Commission	y	y	n	y	y	y	n	y	y	y	y	y	y	y	n	y
Commission can intervene before decisions are adopted/finalized	y	y	n	y	y	y	n	y	y	y	y	y	y	y	n	y
Member of horizontal network	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y

Note: Not all criteria listed here are considered in the body of the text.

Table A.3.: **Independence items**

Regulator	IBPT	CREG	SRTF	CC	ARCEP	CRE	ARAF	AdC	BNetzA	BKartA	OPTA	NMa	Ofcom	Ofgem	ORR	OFT
Formal independence requirement	y	y	n	y	y	y	y	y	y	y	y	y	y	y	y	y
Political incompatibility provisions	y	y	n	n*	y	y	y	y	y	y	y	y	y	y	y	y
Exclusive responsibilities and tasks	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
No ministerial approval of decisions	y	y	n	y	y	y	y	n	n†	n†	n†	n†	n	y	y	y
No indirect or direct ministerial instructions/directions	n	y	n	y	y	y	y	y	y	y‡	y	y	y	y	y	y
Independent internal organization and operations	y	y	n	n	y	y	y	y	y	y	y	y	y	y	y	y
Separate annual budget allocation	y	y	n	n	y	y	y	y	y	y	y	y	y	y	y	y
Term specified	6y	6y	6y	6y	6y	6y	6y	5y	5y	n	4y	6y [chair], 4y [others]	n	5y	5y	5y
One term, renewal barred	n	n	n	n	y	y	y	n¶	1x	n.a	n	1 x 4 y	n	n	n	n
Removal only for sickness, incapacity	n	n	n	y	y	n	n	y	n	n	n	n	n	n	n	n

Note: Not all criteria listed here are considered in the body of the text. For multisectoral regulators, coding of legislative provisions was based on provision

* Incompatibility for local, regional political office.

† General instructions possible.

‡ With exception of negative decisions in merger cases.

¶ President renewable once only; no limits on other members

CENTRE ON REGULATION IN EUROPE

Bibliography

Bortolotti, B., Cambini, C., Rondi, L., and Spiegel, Y. (2011). Capital structure and regulation: Do ownership and regulatory independence matter? *Journal of Economics & Management Strategy*, 20(2):517–564.

Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, 33(5):946–967.

Bradley, R. and Terry, M. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Cambini, C. and Rondi, L. (2011). Regulatory independence, investment and political interference: Evidence from eu.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Curtin, D., Mair, P., and Papadopoulos, Y. (2010). Positioning accountability in european governance: an introduction. *West European Politics*, 33(5):929–945.

de Visser, M. (2009). *Network-Based Governance in EC Law*. Hart, Oxford.

de Vries, H. (1995). An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour*, 50(5):1375–1389.

Edwards, G. and Waverman, L. (2006). The effects of public ownership and regulatory independence on regulatory outcomes. *Journal of Regulatory Economics*, 29(1):23–67.

Elgie, R. and McMenamin, I. (2005). Credible commitments, political uncertainty or political complexity? explaining variation in the independence of non-majoritarian institutions in france. *British Journal of Political Science*, 35(3):531–548.

Gilardi, F. (2002). Policy credibility and delegation to independent regulatory agencies: a comparative empirical analysis. *Journal of European Public Policy*, 9(6):873–893.

Gilardi, F. (2005). The formal independence of regulators: a comparison of 17 countries and 7 sectors. *Swiss Political Science Review*, 11(4):139–167.

Gilardi, F. and Servalli, F. (2011). The consequences of regulatory independence: Evidence from bank capital regulation.

Hancher, L. and Larouche, P. (2011). The coming of age of EU regulation of network industries and services of general economic interest. In Craig, P. and De Búrca, G., editors, *The Evolution of EU Law*, volume 2010. Oxford University Press, Oxford.

Hanretty, C. (2010). Explaining the de facto independence of public broadcasters. *British Journal of Political Science*, 40(01):75–89.

Hanretty, C. and Koop, C. (2010). De jure and de facto independence of regulatory agencies. Paper presented at the 2010 Political Studies Association conference.

Hanretty, C. and Koop, C. (2011). Measuring the formal independence of regulatory agencies. *Journal of European Public Policy*, 19(2):198–216.

Hood, C. (1995). Control over bureaucracy: cultural theory and institutional variety. *Journal of Public Policy*, 15(03):207–230.

Kendall, M. and Smith, B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345.

Koop, C. (2011). Measuring and explaining accountability: A comparative study of independent agencies. PhD thesis, European University Institute.

Kydland, F. E. and Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3):473–492.

Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159.

Larouche, P. and De Visser, M. (2006). The triangular relationship between the commission, NRAs and national courts revisited. *Communications and Strategies*, 64:125.

Li, Y. and Waddams, C. (2011). Effect of regulatory reform of mobile telecommunications on efficiency. Unpublished working paper.

Maggetti, M. (2007). De facto independence after delegation: A fuzzy-set analysis. *Regulation & Governance*, 1(4):271–294.

Maggetti, M. (2010). Legitimacy and accountability of independent regulatory agencies: A critical review. *Living Reviews in Democracy*, 2(1).

Majone, G. (1994). The rise of the regulatory state in Europe. *West European Politics*, 17(3):77–101.

Majone, G. (1996). Temporal consistency and policy credibility: Why democracies need non-majoritarian institutions. EUI Working Paper, RSC No 96/57.

Philp, M. (2009). Delimiting democratic accountability. *Political Studies*, 57(1):28–53.

Radaelli, C. (2004). Getting to grips with quality in the diffusion of regulatory impact assessment in Europe. *Public Money and Management*, 24(5):271–276.

Yesilkagit, K. and Van Thiel, S. (2008). Political influence and bureaucratic autonomy. *Public Organization Review*, 8(2):137–153.

CENTRE ON REGULATION IN EUROPE

Centre on Regulation in Europe (CERRE) asbl
Rue de l'Industrie 42 Box 16 - B-1040 Brussels - Belgium
Ph: +32 (0)2 230 83 60 - Fax: +32 (0)2 23 0 83 70 – info@cerre.eu – www.cerre.eu