Measuring the Formal Independence of Regulatory Agencies

Chris Hanretty and Christel Koop

ABSTRACT

Whilst the literature on delegation has discussed at length the benefits of creating independent regulatory agencies (IRAs), not much attention has been paid to the conceptualization and operationalization of agency independence. In this study, we argue that existing attempts to operationalize the formal political independence of IRAs suffer from a number of conceptual and methodological flaws. To address these, we define what we understand by independence, and in particular formal independence from politics. Using new data gathered from 175 IRAs worldwide, we model formal independence as a latent trait. We find that some items commonly used to measure independence – notably, the method used to appoint agency executives and the scope of the agency's competences – are unrelated to formal independence. We close by showing that our revised measure partially changes conclusions about the determinants and consequences of formal independence.

KEY WORDS Formal independence; regulatory agencies; measurement; item response theory.

INTRODUCTION

Over the past thirty years governments have liberalized markets and privatized state-owned enterprises. To regulate these new markets, governments have often created independent regulatory agencies (IRAs). These organizations have spread quickly (Jordana et al. 2009; Gilardi 2005b), their independence justified as a solution to politicians' time-inconsistent preferences (Kydland and Prescott 1977); a way of introducing technocratic Pareto-improving regulation (Majone 1994), or as a way of prioritizing technical expertise in regulatory policy (Bawn 1995).

To test these claims, political scientists have quantified the degree of independence that IRAs have. In this paper, we argue that previous efforts to operationalize independence – and in particular, the degree of independence found in law – suffer from a number of conceptual and methodological flaws. To remedy these flaws, we model the formal political independence of IRAs as a latent trait, based on new data that we have gathered. This modelling strategy suggests that whilst many of the items are reliable indicators of formal independence, some items and response categories are unrelated to the latent trait, substantiating our theoretical arguments about previous indices. Our revised measure partially changes conclusions about the determinants and consequences of formal independence.

REGULATION AND INDEPENDENCE

Regulation can be carried out by a variety of actors, including IRAs. By regulation, we mean the promulgation and enforcement of an authoritative set of rules governing the private sector, excluding taxation, subsidization, and public ownership, but including rule-creation, the evaluation and scrutiny of economic behavior, and the application of sanctions for non-compliance with rules. IRAs are bodies which regulate, which possess some public authority, and which are not hierarchically subordinate to elected politicians (cf. Thatcher and Stone Sweet

2

2002: 2). We are interested in agencies' independence *from politics*. Agencies can be independent from a range of actors – regulated industries, civil society groups, and the public at large – but we exclusively focus on the independence of regulators from governments, parliaments, parties, and individual politicians. By political independence, we mean the degree to which the day-to-day decisions of regulatory agencies are formed without the interference of politicians and/or consideration of politicians' preferences (cf. Elgie 1998: 55). More particularly, we are interested in agencies' *formal independence* from politics – the degree of independence from politics inherent in those legal instruments which constitute and govern the agency. Formal independence stands in contrast to actual, or de facto independence: the link between the two is an important area of research (Hayo and Voigt 2007; Rosas 2009; Hanretty 2010). Formal independence is worth studying if we are interested in why politicians delegate power: drafting and passing a statute is the ultimate act in delegating power, in both senses. Formal independence is also worth studying if we are interested in regulators' independence in real life. Only by measuring formal independence properly can we test whether actual independence is strongly affected by formal independence, or whether it depends more on national legal-administrative traditions. If formal independence does matter strongly for actual independence, and if more independent agencies deliver better policy, then there is value in studying formal independence.

PREVIOUS INDICES

Existing indices of formal independence suffer from several flaws. These flaws involve the conflation of breadth of powers with degrees of independence, the conflation of the absence of provisions prohibiting an action with permissibility of that action, the assumed order and interval level of the response categories, and the arbitrary weighing of items scores. Table 1 summarizes a number of indices which we cite in our discussion.

TABLE 1

1. Conflating breadth of powers with degree of independence

Many indices, particularly those specific to individual policy fields, conflate breadth of powers with independence. Bluntly, the more things an agency can do, the more independent it is. Thus, Elgie writes that:

> "if the central bank has a large number of monetary policy instruments at its disposal and [...] may use them without restrictions, then the degree of economic independence from the government is high" (1998: 55).

Similarly, Edwards and Waverman state:

> "Greater powers to intervene in interconnection disputes confer greater independence for [*sic*] the NRA [National Regulatory Authority]" (2006: 60).

General purpose indices make similar assumptions. Elgie and McMenamin and Gilardi assume that IRAs which have exclusive competence in their sector – that is, 'more' power in a zero-sum division of power between potential regulators – have greater formal independence. Wonka and Rittberger (2009), who adapt Gilardi's index to the context of the European Union, presume that EU agencies which can take binding decisions are more politically independent than agencies which do not possess this competence. Whilst the breadth of powers an agency may affect its actual independence, and may in turn affect policy outcomes, it is analytically distinct from agencies' formal independence. Formal independence refers to the legal ability of an agency to

make decisions without political interference. The powers or competence of an agency, by contrast, refer to the range of policy instruments the agency has to regulate an industry, and the range of activities which fall under the agency's authority. An agency may possess limited powers but exercise them independently; or it may possess a wide range of powers and exercise them with no independence. By including the number of monetary policy instruments or powers over interconnection issues in indices of independence, one ceases to test the impact of independence on inflation, or interconnection rates. Rather, one tests the impact of having an independent regulator with powers adequate for controlling inflation and interconnection rates respectively. We therefore cannot assess whether it is the adequacy of the regulator's powers, or its independence, which has an impact. Moreover, by conflating powers with independence, we can no longer analyze whether the range of powers which agencies possess has an effect on their degree of formal independence, as demonstrated by Christensen and Nielsen (2010).

2. Conflating the absence of provisions prohibiting an action with the permissibility of that action

Typically, the issues addressed by indices of formal independence are explicitly addressed by legislation. Sometimes, however, certain points – often those relating to term length, re-appointment, and dismissal – are not addressed. The absence of provisions may be significant. Legislation which is verbose may guarantee high formal independence throughout; laconic legislation may guarantee low formal independence even on issues where it is explicit. But this assumption is often built in to many indices, which take the absence of provisions as a low-independence response.

For example, Gilardi (2002) includes an item about incompatibility between membership of the agency board and other government office. If the legislation does not specify an

incompatibility between board membership and other government office, it is scored the same as legislation which states that board membership and government office are compatible. These, however, are not the same. In those countries in which politicians are disposed to grant high formal independence to regulatory agencies, politicians may simply assume that no holder of government office would take up such a position. Counterfactually, had the law addressed this particular issue, it would have specified the incompatibility.

One might circumvent this by assuming that legislation establishing IRAs is intended as a limit on governmental action, and that in interpreting formal provisions one may adopt the maxim that whatever is not explicitly prohibited is permitted. But this will not do. First, this maxim has not been applied, or if it has been applied, has been applied inconsistently. Whilst the index of Cukierman, Webb and Neyapti equates an absence of provisions regarding dismissal with a prohibition of dismissal, Gilardi's and Elgie and McMenamin's indices equate absence of provisions regarding dismissal with the permissibility of dismissal.

Second, this maxim does not apply to legislative lacunae which do not establish permissions and prohibitions – such as provisions on term length. Moreno et al. (2003) assume that the absence of a specified term length is equivalent to an arbitrarily short term. Yet the implicit interpretative maxim – that any gaps in statute should be interpreted to favour the government – is extremely strong, and has not, as far as we are aware, been argued for. Thus, although there is some presumption that legislative lacunae should be associated with lower formal independence, this assumption should be tested instead of being built into the index.

3. Assumed order of response categories

For some items, the order of response categories makes intuitive sense – provisions on term length, for instance, are ordered in terms of their duration, a natural quantity. Yet the majority of

items follow orderings which are only weakly justified. Many response categories, for example, are ordered in terms of distance from the executive. Thus, agencies which take all of their funding from the executive are scored less than agencies which take only some funding from the executive, which are scored less than agencies which take all of their funding from other sources. This ordering 'works' in the case of funding, but for provisions on appointment – where appointment by the legislature is taken to signify more independence than appointment by the executive – the order can easily be challenged. If we are interested in independence from politics *simpliciter*, why is independence from legislators different to independence from cabinet ministers – particularly when the cabinet enjoys a legislative majority? The involvement of the legislature may even damage independence: appointment systems where multiple board members are appointed by parliament by cumulative voting may allow each faction within parliament to appoint at its own ideal point, resulting in extremely partisan boards. Smithey and Ishiyama (2000) circumvent this by talking about the number of actors involved in the appointment process (which has its rationale in veto players theory) but which again requires further elaboration: Does the legislature count as one actor or many? And how should the relative strengths of these actors be assessed? These remarks suggest that the ordering of response categories, particularly in the field of appointments, must either be better justified or tested.

4. Arbitrary weighting of items/nodes

Most indices assign arbitrary weights to index items or nodes. This is most obvious for the index of Cukierman, Webb and Neyapti, where the authors assign weights based on what seemed "most plausible" to them (1992: 361). Others assign nodes equal weighting, arguing that this is the only defensible choice given ignorance about the relevant contribution of items/nodes (Gilardi 2002: 880). However, decisions about how to group index items into index nodes run the risk of

reintroducing arbitrariness, particularly where the number of items for each node varies considerably.

As far as we know, no index of independence has assigned weights to items on the basis of a reasoned assessment of the items' relative contribution to independence. Measures based on principal components analysis of the agency-item matrix are closest in spirit (Gual and Trillas 2006; Edwards and Waverman 2006), implicitly assigning greatest weight to those items which have the highest loading (Banaian et al. 1998). This method of weighting, however, treats category response scores as interval level data, which we argue is incorrect.

5. Assumed interval level of item responses

For items with multiple responses, response scores have typically been equally distributed along the interval [0,1]. For an item dealing with re-appointment provisions, the three possible responses (re-appointment not permitted, re-appointment permitted once, re-appointment permitted more than once) are scored 1, 0.5, and 0 respectively.

Absent information about the distance between the different response categories, this is indeed the best option. Nevertheless, there are good reasons to assume that such options are not equally spaced. Permitting re-appointment is assumed to damage independence insofar as principals can use the threat of non-reappointment to coerce their agents. But the value of re-appointment to existing agents may be reduced if they have already served two terms (they may wish to retire or move to another position); so the damage to independence from permitting multiple re-appointment might not be that much greater than the damage from permitting re-appointment at all. A measure of independence which recognized this would represent an improvement.

MODEL

We have criticized existing indices which specify, ex ante, the weight of each item, and the score associated with each item response. We seek to improve on these indices by using data to estimate the weights and scores of items and item responses respectively on a single, latent, trait; and to exclude items for which reasonable weights and scores cannot be estimated.[1] There are a variety of techniques for reducing multiple manifest indicators to a single latent trait. One technique uses item response theory (IRT). IRT has become increasingly popular, and can be used in any situation where multiple respondents give responses to a common battery of items. IRT is most often used in educational testing, where respondents are test-takers, items are exam questions, and the response categories involve a choice between multiple options, one of which is the 'correct' answer. In this scenario, the response is essentially dichotomous: the respondent gets the answer right or wrong. The latent trait is also simple: respondents who get more, difficult questions correct have greater subject-specific knowledge, or greater intelligence, or both. By using IRT instead of summing the number of correct responses, researchers can find out:

- which questions are more difficult than others (formally: require more of the latent trait; have a higher threshold parameter);
- which questions do not map on to the latent trait (formally: have a small or negative discrimination parameter);
- which respondents are more knowledgeable (formally: have more of the latent trait)

We extend this model to our present situation: the respondents are the agencies, the items are the items we have previously discussed from the Gilardi index, and the latent trait is formal independence. Our response categories are polytomous, not dichotomous: the categories are

ordered in terms of the (increasing) amount of the latent trait required to answer in each category. Thus, we stipulate that appointing board members for six years requires more of the latent trait than appointment for five years, which requires more than four years, and so on. Each of these response categories is then given its own threshold parameter, allowing us to work out how much more difficult each response category is when compared to the previous response category. Instead of scoring the interval between each response category evenly, we can instead test whether moving to appointment for six years represents a bigger leap for formal independence than the leap from four to five years.

The model we use is presented formally in the Appendix. Many of the points we demonstrate using this model could also be demonstrated by using other techniques, such as factor analysis or principal components analysis (Banaian et al. 1998). IRT is related to confirmatory factor analysis, but the "IRT perspective posits a nonlinear relationship between the underlying/latent construct and the observed score at the item/subscale level [w]hereas the [confirmatory factor analysis] perspective… assumes a linear relationship at the item/subscale level" (Raju et al. 2002: 520). If we were to use confirmatory factor analysis, we would ignore our own warning about treating item responses as interval-level data. Nevertheless, many of the terms we use can be understood in terms of their factor analytic analogues, and readers may wish to read 'factor loading' for 'discrimination parameter'.

In what follows, we plot the discrimination and threshold parameters for items and item responses. This model allows us to answer a number of substantive questions about indices of formal independence and the item of which they are composed. First, it enables us to tackle the assumed interval level of the response categories. Existing indices assume that response categories are equally spaced: there is no such restriction in this model. Second, the model allows us to deal with the problem of item weights, by letting the data establish the contribution of each

item. The item discrimination parameter (analogous to the factor loading) not only tells us which items discriminate between subjects of otherwise similar abilities, it also gives the weighting each item would have were one to calculate the factor score for each correspondent. Thus, an item with a low item discrimination parameter is equivalent to an item with low weight in a more traditional index. Third, the model allows us to tackle indirectly the ordering of response categories. Items for which the response categories are misordered are likely to result in either a negative discrimination parameter (suggesting the ordering should be reversed), or a negligibly small discrimination parameter.

DATA

To test these ideas about index creation and the independence of regulatory agencies, we gathered information on the legal provisions governing regulatory agencies worldwide. We constructed a population of regulatory agencies in seven sectors – competition, financial markets, telecommunications, energy, pharmaceuticals, food safety, and the environment (cf. Gilardi 2005a) – by taking the membership lists of the international peak organizations in these fields,[2] resulting in a population of 502 organizations.

We surveyed these organizations on-line (see Hanretty and Koop 2009).[3] Although at several points during the survey we reminded participants to answer "thinking just about the features written in the law", there is no guarantee that respondents' answers are correct statements of the law; some error is inevitable. To minimize the likelihood of respondents answering in a socially-desirable, pro-independence fashion, the survey made no reference to 'independence', only to the governance of the regulatory agency. To test reliability, we analyzed the statutes of the twenty European competition authorities which filled in the survey. Levels of agreement between the two

sets of data were high: the value of Krippendorff's alpha for all 616 non-missing responses was 0.79.[4] We asked respondents about all of the provisions used in the Gilardi index; except that we asked separately whether there was specific provision in the legislation and we separated this response out.[5]

ANALYSIS

FIGURE 1

Figure 1 shows the results of our initial model, including all items.[6] The figure plots the threshold parameters for each response category and each item respectively. Each filled point shows the threshold for a dichotomous item, which is the point at which the respondent has just enough of the latent trait to have a 50% probability of answering in either the higher or lower response category. For example: an agency with a level of formal independence roughly equal to one is indifferent between having a formal statement of independence (*indepagency*) and going without. Numbered points show the threshold between items with multiple response categories. Thus, the point marked '1' is the point at which the regulator has a 50% probability of answering in either the zero-th or the first response category. For example: an agency with a level of formal independence roughly equal to two is indifferent between having its internal organization (*Intorg*) decided by the executive only (the zero-th response category), and having its internal organization decided by the executive and the agency jointly.

The right column lists the discrimination parameters (β) with 90% credible intervals. If items were equally useful in discriminating between agencies with similar levels of formal independence, then all discrimination parameters would be equal and positive. This is not the

case: indeed four items have discrimination parameters which are close to zero or negative. This means that higher response categories are associated with lower levels of formal independence, not less. So, greater reporting obligations to the legislature are associated with more formal independence, not less. The same is true for appointment – whether of board members or chief executives – by the executive. Finally, the discrimination parameter for exclusive competence is close to zero. These negative or near-zero discrimination parameters suggest that these four items are tapping something other than formal independence. We consequently removed these items and re-estimated the model. The results are shown in Figure 2.

FIGURE 2

The estimates enable us to address the practical and conceptual issues raised above. Most obviously, the figure allows us to say which provisions indicate low or high formal independence. Starting from the perspective of a politician reluctantly prepared to grant formal independence to an agency, we can see that moving from budget control by the executive or the legislature, to budget control by an audit office, is a relatively easy concession – perhaps because audit offices are themselves agents of politicians. Similarly, moving from unlimited re-appointment to limited re-appointment is another easy concession, though it is a concession in an area which 'matters more' (has a higher discrimination parameter) for the formal independence of the agency. Conversely, if we start from the perspective of a politician who wishes to grant the maximum amount of formal independence to an agency, then moves such as granting the agency complete control over its internal organization, or appointing agency heads for terms of greater than six years, are the most significant moves – and the hardest concessions for the reluctant politician to make. Finally, if we start from the perspective of the time-pressed researcher, then if we had

information about only one aspect of the agency's formal independence, and wished to infer its overall independence from that datum, we should ask about the term of board members, since this item has a large discrimination parameter, and spans a wide range of the latent trait.

We turn now to the conceptual issues we discussed. First, we argued that conflating breadth of powers with independence was misguided. The negligible discrimination parameter associated with our 'exclusive competence' item showed that this criticism is substantiated by the data. Second, we argued that it was not right to conflate the absence of a provision prohibiting an action, with the permissibility of that action. Our model tests for this by allowing the parameter for having a provision to vary. Where the threshold for having any provisions is to the left of the lowest threshold for any specific provision, then the absence of provisions does seem to indicate low formal independence. This is true for provisions on agency head and agency board member term length, as well as provisions for whether or not the head of the agency may hold other posts in the public administration. It is not true for provisions on dismissal and board incompatibility. This does not mean that having no explicit provisions for dismissal is more indicative of independence than having some provisions at all. Rather, it means that having no explicit provisions on dismissal is compatible with having a higher level of formal independence than agencies with explicit but weak provisions on dismissal. Thus, the lack of provisions cannot be interpreted as a 'low-independence' response.

Thirdly, we argued that some items assumed a questionable order of response categories – and that the categories for items on appointment were more questionable than most. We suggest the negative discrimination parameters for items on appointment indicate that these response categories are misordered, supporting our earlier caution.

Fourthly, we noted that most index items were weighted arbitrarily. Our model resolves this problem. By comparison with the weights used in Gilardi (2002), items relating to the agency

head and board have relatively high weights; conversely items relating to the running of the organization have less weight.

Fifthly and finally, we argued that assuming response categories were interval level (though the best solution given rational ignorance about the contribution of each item) was implausible. Our model bears this out. Natural quantities, such as term length, have item thresholds which are spaced evenly along the latent trait; by contrast, item response categories which have fine gradations, such as provisions on agency over-rule, incompatibility provisions, or provisions regarding funding, are spaced close together and unevenly. Some of these findings – such as the findings about the relative distance between response categories – are subtle, and can only be accounted for by items for ordinal responses. Other more substantively important findings – such as the irrelevance of the appointing actor for formal independence – are more dramatic, but would also be supported by factor analytic techniques.

CONSEQUENCES OF REVISED MEASUREMENT

So far, we have demonstrated that there are good theoretical reasons for dropping certain items from indices of formal independence, and that new data on the formal independence of regulatory agencies shows that these theoretical concerns are not idle: some index items are unrelated to a single unidimensional latent trait of formal independence. In this section, we show that our measurement also makes a difference in analyses of causes and consequences of independence.

Gilardi (2005a) argues that regulators' formal independence is a positive function of (a) whether or not the regulator regulates a privatized utility; (b) whether or not the regulator regulates financial or other market operations; (c) the number of veto players in the polity. Independence is also a quadratic function of government replacement risk: up to a certain point the more frequently governments are replaced by alternate governments with different partisan

compositions, the greater the level of formal independence; after that point, formal independence decreases, as governments expect to return to power quickly and thus see no reason to insulate policy-making.

TABLE 2

To test our measurement, we used Gilardi's raw data as an input to our model, and used the resulting output as a dependent variable in a series of regressions (Models 2 and 3 in Table 2), comparing them to a regression using Gilardi's original measurement (Model 1). Using our measurement, neither of the replacement risk variables is found to be significant, even with Huber-White country-clustered standard errors. The reason for this lack of statistical significance is shown in the third model, which drops the square of replacement risk. Replacement risk returns as a significant variable, but the formal independence of regulatory agencies is now a simple linear function of replacement risk: countries with extremely high levels of replacement risk do not start cutting back on the amount of formal independence they grant their regulatory agencies. In any case, replacement risk has a more modest effect in comparison with the original model; the impact of an increase in veto players, by contrast, is greater.

Second, we can show that our measure is also (very slightly) better related to measures of de facto independence. Maggetti (2007) assesses the degree of independence of sixteen West European regulatory agencies, distinguishing between de facto independence from politicians and de facto independence from those who are regulated. Our measure of formal independence is better correlated with his assessments of actual independence than the previous measure (Pearson's r=0.34 as compared to r=0.23 for the alternate measure).[7] Whilst there may be good reasons why one might not find a positive association between formal and de facto independence,

16

there is a strong presumption that higher levels of formal independence should lead to higher levels of de facto independence, other things being equal. That our measure does demonstrate such a link gives indirect support for our removing some index items we judged, on theoretical grounds, to be irrelevant to formal independence.

CONCLUSIONS

In this article we have defined what we mean by formal independence, and have addressed the flaws present in existing attempts to operationalize this concept. We have modelled new data gathered from regulatory agencies around the world, and our conceptual and methodological criticisms have been borne out – those items which had little conceptual rationale also fit badly our latent trait model of formal independence.

If we are correct, our findings imply that appointment method should be dropped from indices of independence. This is an important finding, since the choice of appointment method features in every index of independence we are aware of. If appointment provisions are unrelated to formal independence, why have they been treated as if they were? One possibility is that whilst the appointment method is unrelated to formal independence, it may be related to actual independence. Nevertheless, future research should assess the exact relationship between the appointment method and formal and actual independence.

We have attempted to model independence as a latent trait to improve measurement. We suggest that this is important for those who use the formal independence of regulatory agencies as a dependent variable – better measures of dependent variables will lead to smaller standard errors – but crucial for those who use the formal independence of regulatory agencies as an independent variable, for here measurement error will bias the coefficients obtained.

A first-best solution to problems of measurement error would be for researchers in this field to employ latent trait models whenever dealing with formal independence. Doing so will enable a check of the goodness of fit both of the overall model and of individual items: whilst we are reasonably confident that the items we have identified as performing poorly are not an artifact of our data, the same items may either perform better in other contexts, or may be joined by additional non-performing items. Such models will also allow researchers to better handle missing data – something which is more important if, as we suggest, the absence of provisions should be treated as a form of missing data, rather than as an indication of low formal independence.

Biographical notes: Chris Hanretty is Lecturer in Politics at the University of East Anglia. Christel Koop is a Researcher at the European University Institute.

Addresses for correspondence: Chris Hanretty, University of East Anglia, Social, Political and International Studies, Norwich NR4 7TJ, United Kingdom. E-mail: c.hanretty@uea.ac.uk / Christel Koop, European University Institute, Department of Political and Social Sciences, Via dei Roccettini 9, 50014 San Domenico di Fiesole, Italy. E-mail: christina.koop@eui.eu.

ACKNOWLEDGEMENTS

APPENDIX

The model we use is Samejima's Graded Response Model (Samejima 1970), which is equivalent to the mixed-response factor analysis model proposed by Quinn (2004) and implemented in MCMCpack (Martin et al. 2009). Formally, we have i regulators who respond to $j = 1...J$ items, each item having a number of response categories C_j. Each regulator has an amount of the latent trait θ_i; each item has a difficulty parameter, or threshold (α_j) and a discrimination parameter (β_j). The *latent response* of each regulator to each item (x^*_{ij}) is therefore

$$x^*_{ij} = \alpha_j + \beta_j\,\theta_i + \varepsilon_{ij}$$

with $i = 1,...,N$, $j = 1,...,J$ and errors distributed normally. This latent response then becomes a manifest response as it is discretized by a number of thresholds for each item, γ_{jc}, for $j = 1...J$ and $c = 0...Cj$, with $\gamma_{j0} \equiv -\infty$, $\gamma_j c_j \equiv \infty$, and $\gamma_{j1} \equiv 0$ for the purposes of identification:

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

The more formal 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.

NOTES

1 This approach can also be challenged: our theory might suggest that some items ought to be included regardless of whether they are empirically useful; or that some items ought to have a certain score based on our prior knowledge. But we believe that our prior knowledge about independence is not sufficiently strong to permit these kinds of judgements.

2 For competition, the International Competition Network; for financial markets, the International Organization of Securities Commissions; for telecommunications, the International Telecommunication Union; for energy, the International Energy Regulation Network; for pharmaceuticals, the International Conference of Drug Regulatory Authorities; for food safety, the Global forum of food safety regulators; for the environment, the European Union Network for the Implementation and Enforcement of Environmental Law.

3 We sent the surveys to all 706 contact points on the membership lists of the 502 peak organizations. These lists varied considerably in detail, and we often were not able to say which post within the organisation our respondents held. Nevertheless, the fact that initial respondents forwarded on our questionnaire to other individuals within the organisation, and the fact that we received hardly any duplicate surveys – even though we sent the survey to more contact points than organizations – suggests that some coordination took place within the organizations.

4 We have assumed ordinal responses. Missing responses resulted from inapplicable questions (questions concerning the board for single-headed authorities) and from non-response. The level of agreement would have been higher had it not been for our cautious interpretation of country-specific statutory provisions. We were often not able to determine whether nominations were binding, and hence opted for the response category "No specific provisions", whilst the respondents could usually give a definite answer to this question.

5 The survey is not a random sample of the population; real GDP per capita was a statistically significant predictor of responding to the survey. Our results about the structure of formal independence may not therefore be generalizable to least developed or developing countries. Response rate was not, however, significantly affected by the language of the survey, or the type

of legal system. Our results can therefore be generalized across legal systems. Descriptions of each item and associated response categories, as well as response frequencies, can be found at http://chrishanretty.co.uk/

6 These results obtained after running MCMCpack's MCMCordfactanal routine for 1 million iterations after a 100,000 iteration burnin, thinned every 200 iterations. Inspection of trace plots for the α, β and γ parameters showed no problems with convergence. Similar discrimination parameters were obtained using marginal maximum likelihood as implemented in the ltm package (Rizopoulos 2006), and through normal factor analysis imputing missing values with the item mean, and taking the first unrotated factor.

7 In neither case is the correlation significant, which is mainly a consequence of the low number of cases (N = 16).

REFERENCES

Banaian, K., Burdekin, R. and Willett, T. (1998) 'Reconsidering the principal components of central bank independence: The more the merrier?', *Public Choice* 97(1): 1-12.

Bawn, K. (1995) 'Political control versus expertise: Congressional choices about administrative procedures', *American Political Science Review* 89(1): 62-73.

Christensen, J. G. and Nielsen, V. L. (2010) 'Administrative capacity, structural choice and the creation of EU agencies', *Journal of European Public Policy* 17(2): 176-204.

Cukierman, A., Webb, S. and Neyapti, B. (1992) 'Measuring the independence of central banks and its effect on policy outcomes,' *World Bank Economic Review* 6(3): 353-398.

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. (1998) 'Democratic accountability and central bank independence: Historical and contemporary, national and European perspectives', *West European Politics* 21(3): 53-76.

Elgie, R. and McMenamin, I. (2005) 'Credible commitment, political uncertainty or policy complexity? Explaining variations 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. (2005a) 'The formal independence of regulators: A comparison of 17 countries and 7 sectors', *Swiss Political Science Review* 11(4): 139-167.

Gilardi, F. (2005b) 'The institutional foundations of regulatory capitalism: The diffusion of independent regulatory agencies in Western Europe', *The Annals of the American Academy of Political and Social Science* 598(1): 84-101.

Gual, J. and Trillas, F. (2006) 'Telecommunications Policies: Measurement and Determinants', *Review of Network Economics* 5(2): 249-272.

Hanretty, C. (2010) 'Explaining the de facto independence of public broadcasters', *British Journal of Political Science* 40(1): 75-89.

Hanretty, C. and Koop, C. (2009) 'Compararing regulatory agencies: Report on the results of a worldwide survey. EUI Working Paper RSCAS 2009/63. Available at: http://cadmus.eui.eu/bitstream/handle/1814/12877/RSCAS_2009_63.pdf?sequence=1

Hayo, B. and Voigt, S. (2007) 'Explaining de facto judicial independence', *International Review of Law and Economics* 27(3): 269-290.

Jordana, J., Levi-Faur, D. and Fernandez, X. (2009) 'The global diffusion of regulatory agencies and the restructuring of the state', Available at:. http://law.huji.ac.il/upload/Global_Diffusion__23_December_2008.pdf

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

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

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

Martin, A., Quinn, K. and Park, J. H. (2009) MCMCpack: Markov chain Monte Carlo (MCMC) Package.

Moreno, E., Crisp, B. and Shugart, M. (2003) 'The accountability deficit in Latin America', in S. Mainwaring and C. Welna (eds.), *Democratic Accountability in Latin America,* Oxford: Oxford University Press, pp. 79-132.

Quinn, K. (2004) 'Bayesian factor analysis for mixed ordinal and continuous responses', *Political Analysis* 12(4): 338-53.

Raju, N., Laffitte, L. and Byrne, B. (2002) 'Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory', *Journal of Applied Psychology* 87(3): 517-529.

Rizopoulos, D. (2006) 'ltm: An R package for latent variable modeling and item response theory analyses', *Journal of Statistical Software* 17(5): 1-25.

Rosas, G. (2009) 'Trust in elections and the institutional design of electoral authorities: Evidence from Latin America', *Electoral Studies* 29(1): 74-90.

Samejima, F. (1970) 'Estimation of latent ability using a response pattern of graded scores', *Psychometrika* 35(1): 139-139.

Smithey, S. and Ishiyama, J. (2000) 'Judicious choices: Designing courts in post-communist politics', *Communist and Post-Communist Studies* 33(2): 163-182.

Thatcher, M. and Stone Sweet, A. (2002) 'Theory and practice of delegation to non-majoritarian institutions', *West European Politics* 25(1): 1-22.

Wonka, A. and Rittberger, B. (2009) 'How independent are EU agencies?', RECON Online Working Paper 2009/12. Available at:

http://www.reconproject.eu/main.php/RECON_wp_0912.pdf?fileitem=3555641.

Table 1: Indices of statutory independence

| Reference	Independence of what from whom?	Clusters/nodes	Non-exhaustive list of items	Item response type	Aggregation method
Gilardi (2002)	IRAs from politicians	Agency head (6); Agency board (6); Relationship with gov't, parl't. (4); Financial, organizational autonomy (4); Exclusive competence (1)	Agency head cluster include items on term length, appointing body, dismissal, incompatibility, renewability, and requirement for independence	Dichotomous: 3; polytomous: 18; median number of response categories 4	Mean of node scores
Elgie and McMenamin (2002)	Agencies from politicians	Agency head (4); Agency board (4); Regulatory competencies (5)	Agency head and Agency board cluster as per Gilardi, less incompatibility and independence items	Dichotomous: 5 (all relating to competences); Polytomous: 8	Mean of appointments node scores, plus competencies node score, divided by two
Cukierman et al (1992)	Central banks from governments	Central bank head (4); Regulatory competences vis-a-vis govt (3); Central bank objectives (1); Limitations on lending (8)	Agency head cluster as per Gilardi, less renewability and independence items	Dichotomous: 2; Polytomous: 14; median number of response categories 4	Arbitrarily-weighted mean
Smithey and Ishiyama (2002)	Courts from politicians	None – flat structure (6)	No. actors involved in appointment; term length, dismissal provisions; org. autonomy, presence of a priori review; decisions irreversible or not	Dichotomous: 2; Polytomous: 4; median number of response categories 3	Mean of item scores
Edwards and Waverman (2006)	Telecoms regulators from governments	None – flat structure (12)	Variety of indicators, including whether regulator is multi-sectoral, multi-member, has exclusive competence, and has power over interconnect rates	Dichotomous: 8; Polytomous: 4; median number of response categories 3	Mean of item scores

Figure 1

26

Figure 2

27
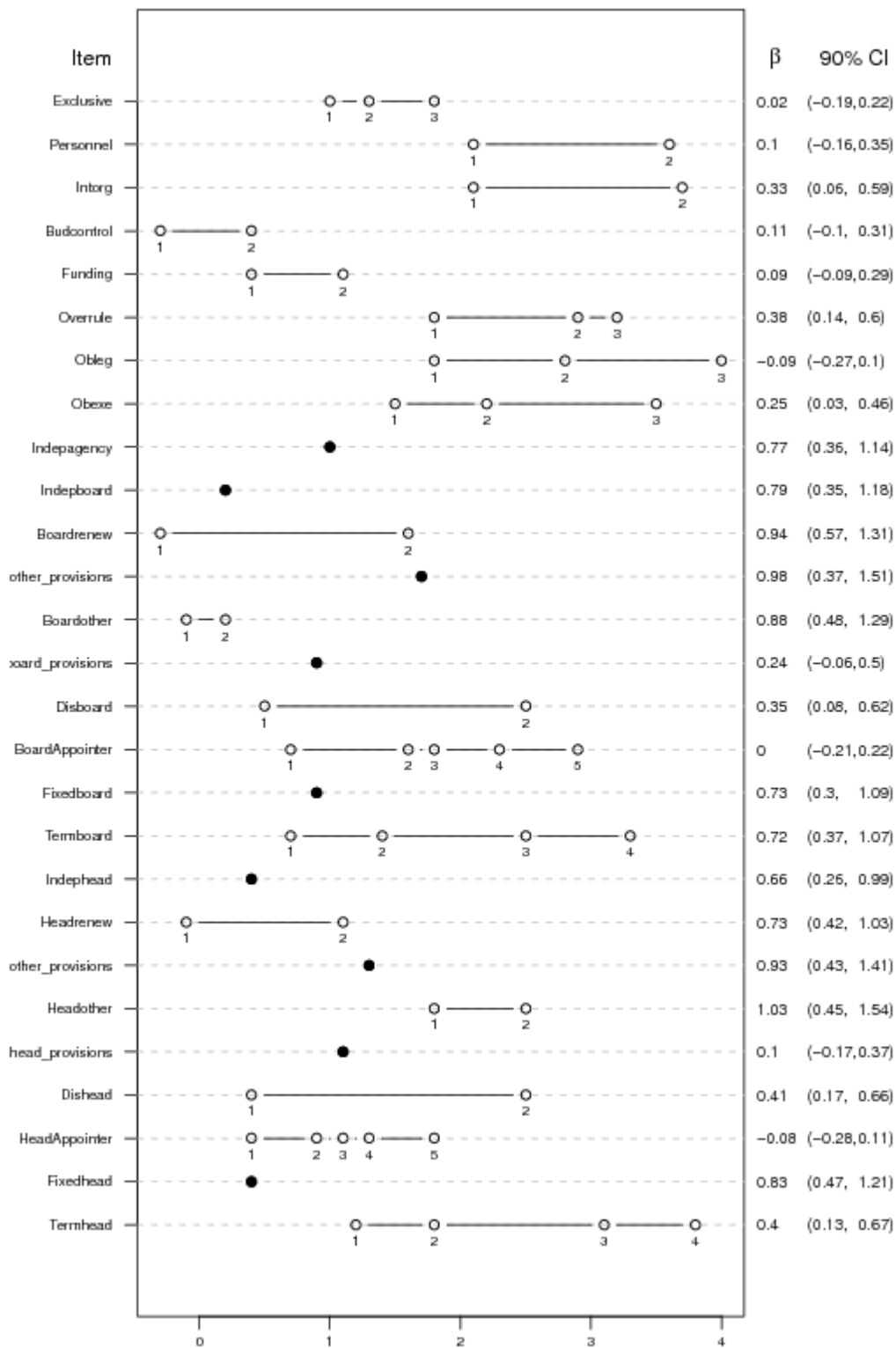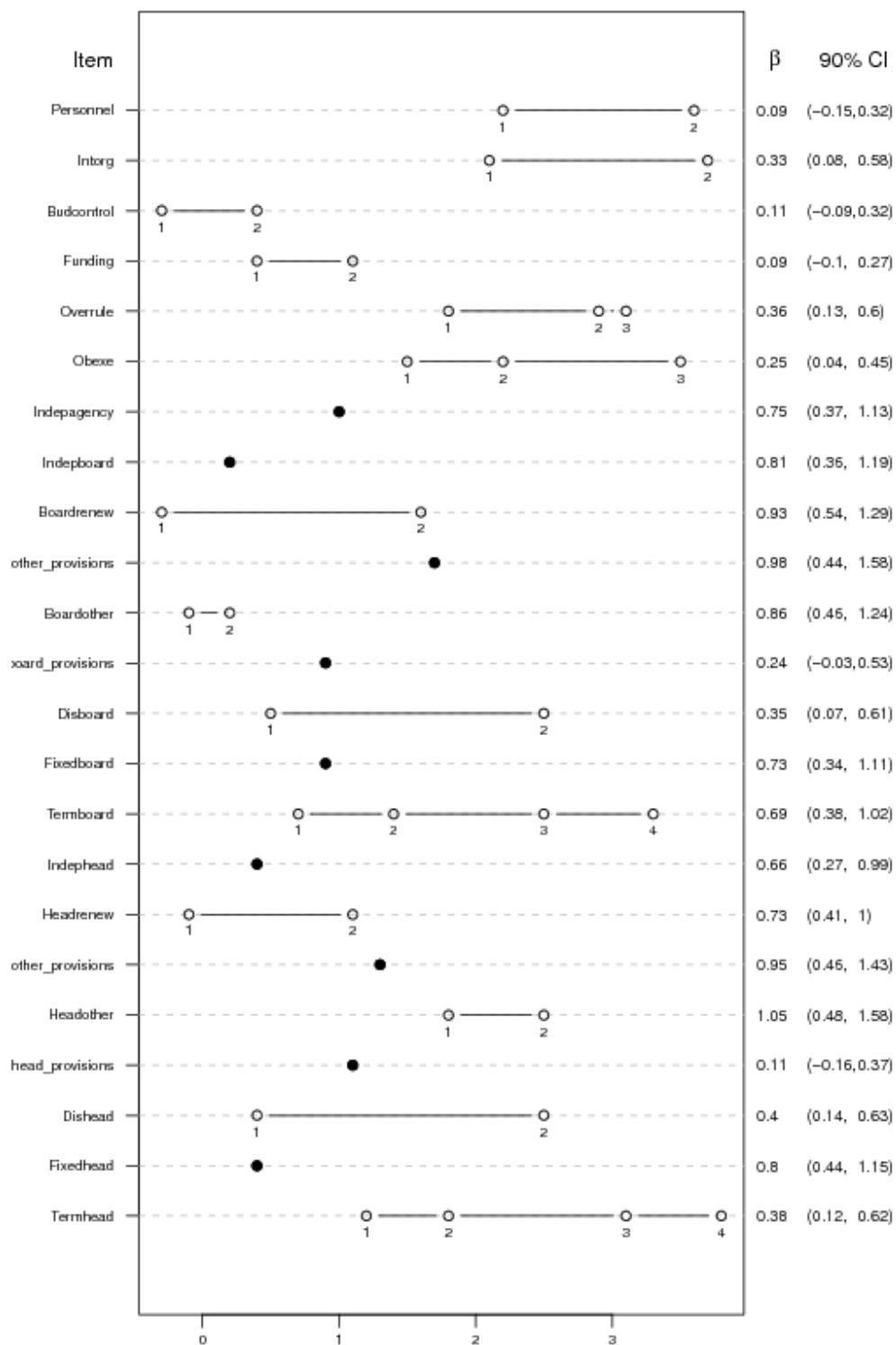
Table 2: Models of delegation

	Original measure	Revised measure	Revised measure, reduced model
Intercept	0.05	0.11	0.08
	(0.06)	(0.07)	(0.06)
Utilities	0.32*	0.29	0.29*
	(0.05)	(0.07)	(0.07)
Financial markets	0.22*	0.23*	0.23*
	(0.04)	(0.05)	(0.05)
Replacement risk	1.49*	0.21	0.53*
	(0.48)	(0.55)	(0.18)
Replacement risk²	-1.89*	0.51	
	(0.88)	(0.94)	
Veto players	-0.06*	-0.06*	-0.07*
	(0.03)	(0.03)	(0.03)
N	106	106	106
R²	0.41	0.37	0.37
Adj. R²	0.38	0.34	0.35
Resid. sd.	0.20	0.21	0.21

Robust standard errors in parentheses. * indicates significance at $p < 0.05$