Measuring portfolio salience using paired comparisons
A new application of the Bradley-Terry model*

Mariana Batista
Assistant Professor/Professora Adjunta
Federal University of Pernambuco
mariana.bsilva@gmail.com

Timothy Power
Professor
University of Oxford
timothy.power@lac.ox.ac.uk

Cesar Zucco
Associate Professor
Getúlio Vargas Foundation
cesar.zucco@fgv.br

March 26, 2018

Abstract
We propose measuring portfolio salience by analyzing paired comparisons data using the Bradley-Terry model. Paired-comparison data are easy to collect using surveys that are user-friendly, rapid, and inexpensive, and we implement the approach with both politicians and experts in Brazil. We show that practitioners and academics have broadly similar views of the relative worth of cabinet posts, and that estimates are robust to variations in implementation of the method, suggesting that either group could be surveyed efficiently to generate measures of portfolio salience. Substantively, we find that respondent valuations of cabinet portfolios deviate considerably from what would be predicted by objective institutional measures such as budget, policy influence, and employment opportunities for political appointees.

Keywords: Portfolio Salience; Coalition Governments; Bradley-Terry Model

*We thank the Brazilian Political Science Association for allowing us to use their member mailing list in the expert survey reported in this paper, Manuella Cattelan and Heather Turner for advice on estimating models within the Bradley-Terry framework, as well as Octavio Amorim Neto and participants in the Workshop “Rethinking Brazilian Politics: New Directions in Research” held at the University of Oxford in 2018. Authors acknowledge support from Getulio Vargas Foundation’s Rede de Pesquisa Aplicada for the 8th wave of the Brazilian Legislative Survey.
The formation of interparty cabinet coalitions is a longstanding theme in the study of parliamentary democracy, and since the Third Wave of democratization it is now a central issue in comparative presidentialism as well. Formateurs have a prize to offer in the government formation process, and this usually takes the form of cabinet positions. The control of specific ministries is important to coalition partners because it represents the opportunity to influence government decisions, so the question becomes one of who gets what.

Scholars of government formation in both parliamentary and presidential systems generally agree that not all cabinet-status positions are of equal importance. Yet for both researchers and formateurs, a proper understanding of coalitional politics invites us to assess the degree of satisfaction of political parties when they are awarded various positions in multiparty cabinets. This is the basic motivation behind the long-running research programme on “portfolio salience,” i.e. the size of the prize. Measuring salience has proved vexing. It is not obvious whether one should use reputational prestige, policy-related measures, or some observed characteristic of the ministry in question (e.g. the overall budget share, the “discretionary” budget, the number of employees, or the management of government contracts). It is further the case that ministries can be important for different reasons. Some of the portfolios responsible for many core policy decisions have, in fact, small budgets or few personnel, thus rendering comparisons based on observed characteristics of the ministries (Mauerberg Junior 2016, Batista 2017), or on revealed preferences of parties (Adachi & Watanabe 2007, Browne & Feste 1975) potentially misleading.

Following seminal work by Laver & Hunt (1992), several variations on expert or elite surveys have been used to gauge portfolio salience (e.g. Kato & Laver 1998, Warwick & Druckman 2001, Druckman & Warwick 2005, Druckman & Roberts 2008). These survey approaches, however, are typically very demanding for respondents. This can be particularly burdensome in the context of governments with many ministries, or where the lineup of cabinet positions changes frequently. In these contexts, having experts rate all the extant ministries is impractical, and obtaining the cooperation of a large number of elected politicians is never easy.

These earlier efforts inspire the present research note, in which we propose the use of the Bradley-Terry model to analyze pairwise comparisons of cabinet positions. The basic Bradley-Terry model is a relatively simple statistical tool that has been applied in many fields, from psychology (to recover “ability” scales from the outcomes of observed pairwise comparisons) to sports journalism (as in the college football team rankings popular in the U.S.). Although the

1For example, the New York Times computer ranking of NCAA football teams recorded three variables about
model can accommodate a number of refinements, its basic variant is enough, as we show, to make both expert and elite surveys feasible. All that is required from respondents is that they pick the more valuable portfolio from a few different pairs of options presented to them.

The approach is very general and could be applied anywhere, but we test it on new data from Brazilian politicians and experts. First, contemporary Brazil has a very large and heterogeneous set of cabinet-level posts. Second, as a paradigmatic case of “coalitional presidentialism” (Chaisty, Cheeseman & Power 2018), Brazil has spawned what is probably the largest literature on the forming and functioning of coalition governments outside of parliamentary democracies (e.g. Abranches 1988, Amorim Neto 2000, Amorim Neto 2006, Zucco Jr. 2009, Power 2010, Raile, Pereira & Power 2011). Our measures of portfolio salience are intended to improve measurement of cabinet coalescence in that literature (Amorim Neto 2000, Mauerberg Junior 2016, Batista 2017).

We present the model and retrieve estimates from a dedicated survey of experts in Brazil, and we compare the results with a similar, smaller survey with politicians. The results of the two surveys are impressively similar and are quite stable to variations in the respective datasets.

1 Measuring Portfolio Salience Using the Bradley-Terry Model

The BT model, is a ranking model based on contests between alternatives. The attributes may be based on subjective evaluations of properties of the objects made by referees or on “objective” outcomes such as estimating the strength of football teams from game results, or the quality of scientific journals by citation patterns.

The model was originally proposed in Bradley & Terry (1952), and has been extended in several ways over the years in many different fields. Some closely related models have been thoroughly employed in the political science literature. Multinomial choice models and variations thereof developed for specific applications (e.g. Groseclose & Stewart III 1998, Desposato 2009), for instance, have been widely used, and articles employing conjoint analyses implicitly rely on models that are very similar to, if not identical to, variations of the BT model. Direct application every matchup: who won, by what margin of victory, and against what quality of opposition. Estimation of the model originally required the use of an IBM 370 mainframe (New York Times 1979).

Groseclose & Stewart III (1998) estimate the implicit worth of each committee assignment in the US Congress from observed transfers between committees and Desposato (2009) to estimates the implicit worth of parties from party switching data. However, in these contexts, a move from one committee or party to another meant not only that the receiving entity was preferred over the sending committee or party, but that it was preferred over all other options. Hence, our present problem is simpler, as it is just one of binary choice between two alternatives. As such, we can rely on established versions of the BT model.
plications of the BT model in Political Science, however, are surprisingly quite rare. The BT model was used in an analysis of value orientation in Europe, but the work was conducted by economists and statisticians who generalized the model in important ways to account for missing data and categorical covariates of subjects making the decisions (Francis, Dittrich, Hatzinger & Penn 2002). One noteworthy exception is work by Loewen, Rubenson & Spirling (2012), which employed the BT model to analyze data from a survey experiment designed to assess which types of arguments are most persuasive in a referendum. In this paper, which is the one that most closely resembles ours, respondents were presented with one out of six arguments in favor and one out of six against a proposed change in the electoral system, which would be on the upcoming ballot. There was a relatively small set of items to be ranked (12 in total) and each of the 520 respondents got to make only one pairwise comparison. In our application, in contrast, we have a larger number of items (37 ministries in total), so we opted to have each participant in both surveys rank a larger number of pairwise comparisons.

In applications of the BT model, the items being evaluated are sometimes referred to as ‘players’, those making the decisions are ‘referees’ or ‘subjects’, depending on the application, and the underlying scale being recovered is referred to as the ‘ability’ of the players or ‘worth’ of the items. For consistency with our application, we refer to players as ‘items’ or ‘ministries’ ($M$), ‘referees’ as respondents and ‘ability’ as the value or worth of each item ($\alpha$). In our application, for each ‘contest’ or comparison between two ministries, our subjects were asked to choose the most valuable one.\footnote{Ties were not allowed.} We are not interested in examining the impact of specific attributes of each item on this choice, as is typically the case in conjoint analysis applications. Instead, we simply want to retrieve the underlying relative value for each item.

The BT model is ideally suited for this because it is built on the probability that one item $M_i$ is chosen over another $M_j$ in a pairwise contest, which we represent as $\pi_{ij}$. This probability can then be expressed as function of the underlying worth of each item, which we label $\alpha_i$, such that:

$$\pi_{ij} = \frac{\alpha_i}{\alpha_i + \alpha_j} \quad (1)$$

In its original formulation, this probability of selection is then expressed in terms of exponential functions $\exp(\lambda_i) \equiv \alpha_i$ of the underlying worth and by employing a logit link function. This leads, to:

$$\text{logit}(\pi_{ij}) = \lambda_i - \lambda_j \quad (2)$$
In short, $\lambda_i$ is a transformation of the underlying worth $\alpha_i$, which determines the probability of one item being chosen over another in pairwise contexts, as defined in Equation 3.

This logit specification is precisely what characterizes the BT model. Similar models have been proposed that use alternative specifications. As discussed in previous work on the subject (Luce, Krantz, Suppes & Tversky 1990, Loewen, Rubenson & Spirling 2012), this specification meets the strong stochastic transitivity condition, such that when $\pi_{ij} \geq 1/2$ and $\pi_{jk} \geq 1/2$, then $\pi_{ik} \geq \max \pi_{ij}, \pi_{ik}$. This is important because items that are not directly compared in any contest can be indirectly compared, as long as the design is connected (i.e. that no subset of the items exists that is has not been compared to another subset of the data) (Bradley 1984). Additionally, as long as no item wins all its contests and no subgroup of items exists that its members always win contests against other subgroup of items, the parameters in this model can be estimated by maximum likelihood with either a sum constraint ($\sum_{i=1}^{n} \lambda_i = 0$) or, more commonly, a reference object constraint ($\lambda_i = 0$ for one $i \in \{1, \ldots, n\}$). For the remaining parameters, standard errors are computed as usual. Items that beat items with larger $\alpha$ receive a larger boost to their own $\alpha$ because information from one contest in implicitly used in other contests.

This basic BT setup is often referred to as the “unstructured” model, and it is the workhorse model we employ in this paper. This unstructured model has been extended in many different ways. Some of the most common applications include allowing for the rankings of more than two options, accounting for the order of the items in the pairwise contest (i.e. ‘order effect’ or ‘home-field advantage’), accounting for other contest-specific variables such as experience of the players/items (Stuart-Fox, Firth, Moussalli & Whiting 2006), or even allowing abilities to be modeled as a function of item specific covariates (Springall 1973). In our application, these extensions are not necessary, as we simply want to retrieve the underlying worths of each ministry.

However, we do have to contend with a model-specific form of missing data. Because of the large number of items being rated, not all possible matchups are observed; this is the equivalent to ranking two college football teams that never played each other. By construction of our experimental design, these data are missing at random, so they should not affect estimates directly.

---

4See the appendix for how to get from Equation 3 to 7.

5If, instead, the difference in the worth of each item were modelled using a normal cumulative distribution, it would define the related Thurstone (1927) model (For a detailed discussion, see Bradley 1984, Cattelan 2012).

6Items that show up more often will have smaller confidence intervals.
Ideally, we would also want to account for the fact that results from pairwise comparisons made by the same survey participant are not independent from each other. Although extensions of the BT model that account for referee-specific covariates exist (Böckenholt 2001, Francis, Dittrich & Hatzinger 2010), these are useful only to the extent that one is interested in examining their effect on the outcomes of the comparisons. In our case, we would prefer to account for referee-specific factors through a fixed-effects type of framework. However, this would require that each referee rate all items, as referee effects are only identified in interaction with each item Cattelan (2012).

2 The Data

We fielded two surveys simultaneously in Brazil in 2017, the first with “experts” and the second with incumbent legislators in the National Congress (which we refer to as the “elite” survey). In both, we presented pairs of ministries that existed in 2014 and 2015, which spans the last two full years of president Dilma Rouseff’s administration. A total of 37 ministries were rated, which implies the existence of 666 pairwise comparisons, assuming that the order in which the items are presented to participants is irrelevant. Since it was not feasible to have all participants rate all possible pairwise comparisons, we opted to randomize the items that were presented to each respondent.

The expert survey was conducted with 283 members of the Brazilian Political Science Association (ABCP). The survey—built and distributed through Qualtrics—solicited basic demographic data and asked whether the respondents considered themselves “experts” in executive-legislative relations or other subfields. It then presented respondents with pairwise comparisons of cabinet portfolios. Experts were asked to choose the ministry they thought a typical politician would prefer to obtain for his or her party. Each respondent answered an initial battery of 8 pairwise comparisons, and most opted to undertake 3 additional comparisons. Item non-response was minimal, so we treat it as completely random.

7 Rousseff was elected president in 2010 and reelected in 2014, but was impeached and removed from office in mid-2016, approximately one year prior to our survey.
8 The list of 37 portfolios incorporated traditional ministries as well as some special “secretariats” with cabinet rank. In this research note, we refer to all cabinet-status units as ministries or portfolios. Some portfolios were merged once Michel Temer (Rousseff’s vice president) assumed the presidency, with the number declining first to 30 and then to 25 in late 2016.
9 The question wording varied slightly in the expert and elite versions: “Imagine a hypothetical situation in which a future President of the Republic is sounding out (a political party/your party) about taking on a cabinet-level position in the government. For each one of the pairs below, select the position you think (a typical federal legislator/your party) would prefer.”
Because the BT model uses the transitivity assumption to leverage information from multiple comparisons, we deemed it useful to minimize the number of “lopsided” matchups in the surveys. For this purpose, we stratified the portfolios into four tiers based on our own \textit{a priori} expectations about their reputational value. In 8 of the 11 matchups, the two competing portfolios were drawn from a subset of two adjacent tiers, and six of these involved the highest valued tier. Thus, while portfolios were still randomly drawn, there are more “close” comparisons in our dataset than would have obtained with simple randomization, and also more contests involving high-profile ministries. The survey generated data for a total of 3,101 comparisons. In the expert data, 275 of the possible pairwise 666 contests were never observed, while four contests were observed more than 20 times.

For the elite survey we were further constrained for two reasons. First, the comparisons were embedded in the 8\textsuperscript{th} wave of the Brazilian Legislative Survey (BLS), which included many other items, so space was limited to four pairwise comparisons per respondent. Additionally, the BLS was implemented using both an online version and a paper questionnaire (many legislators prefer the low-tech format). In the electronic version, we implemented a randomization scheme almost identical to the one described for the expert survey, with one comparison being drawn from the entire pool of portfolios, and the others drawn from adjacent tiers. On the hard-copy instrument, we created five different sets of four randomly drawn pairwise comparisons, and the five versions of the paper questionnaire were randomly distributed to incumbent members of Congress.

In both implementations, legislators were asked which ministry they would prefer their party held. A total of 141 legislators participated in the survey, with 62 responding online and 79 on paper. Two respondents failed to answer the pairwise comparison battery and were dropped, but there was no item nonresponse among the remaining legislators. The elite data encompass a total of 556 pairwise comparisons, with 221 different matchups occurring at least once.

In both datasets, all portfolios were rated (at least 102 times in the expert survey and at least 7 times in the elite survey), which allows for an estimation of the worth of all portfolios. Of course, items that appear more often have smaller confidence intervals, and, in general, estimates from the expert dataset are more precise than those from the elite dataset. Substantively, the elite survey is our preferred source of data, as it reflects the perceptions of individuals who operate directly in the arena of government formation. However, in this research note we present both surveys, not only to gauge robustness, but also to assess the feasibility of applying
the BT method to this and other political science measurement tasks.

The survey instrument necessary for pairwise comparisons is very user-friendly for respondents. Several ABCP colleagues reported to us that the survey was “fun,” which may explain why 273 of the 278 participants opted to continue on past the first eight pairwise comparisons. Overall, the median time to completion for the entire expert survey, including the demographic and subjective competence items, was just over 4 minutes. This makes collecting data feasible, even if the number of items to be ranked is, as in our case, quite large.

3 Estimates Values of Ministerial Positions in Brazil

We first present our main results, applying the BT model to data generated by the elite survey. In Table 1, cabinet-level positions are ranked from highest to lowest according to the point estimates of the value of each portfolio ($\alpha$). The first column reports our a priori stratification of the ministries into four tiers. The following columns report our estimates of $\alpha$, which are provided relative to the baseline category (set a posteriori to be the highest-ranked ministry), whose value is arbitrarily fixed at zero. We also present the statistical significance of comparisons of differences in value between each portfolio and this baseline category.

The substantive story is that our initial intuitions were not far off from legislators’ opinions. However, there were some important differences, as exemplified by Ministério das Cidades, which we translate as Urban Affairs. This portfolio was assigned to the second tier by us but was viewed by incumbent legislators as the most valuable ministry. Interestingly, among the top-ranked portfolios, we observe two that can be viewed as important sources of pork and patronage (Urban Affairs and Education), two that are ostensibly “core policy” ministries with few material goods to distribute but considerable policy influence (Finance and Planning), and one that is the quintessential political coordination portfolio (Casa Civil, the presidential chief of staff). This diversity suggests that portfolios may be important for varying reasons that are not easily captured by a common set of observable characteristics, thus reinforcing the importance of a perceptual measure such as the one we propose.

However, it is also the case that our estimates are noisy. The top-ranked portfolio, for instance, cannot be statistically distinguished from the subsequent five in pairwise comparisons.

---

10We used the BradleyTerry2 package for R (Turner & Firth 2012). Other R packages exist, such as Prefmod whose functionality is slightly different, but which generates identical estimates.

11Since 2009, this portfolio has been responsible for the Minha Casa, Minha Vida (My House, My Life) social housing program, one of the largest such initiatives in the world.
More generally, it is not easy to convey the uncertainty about our estimates, as it is not feasible to report $p$-values for all relevant pairwise comparisons in a cogent way. For this reason, we developed and report bootstrapped confidence intervals for the rank-order outcome of each portfolio. We drew 1,000 simulated estimates for each ministry’s position from a multivariate normal distribution with mean and variance-covariance matrix set to the one that was originally estimated. We then computed the rank of each item in each of the simulations, and reported each ministry’s 5th and 95th percentile rank.

These confidence intervals provide a more intuitive way of understanding the uncertainty involved in the estimates. We can say, for instance, that there is a 90% probability that the Ministry of Finance is ranked among the top five portfolios. It is also immediately evident that while there is a clear cluster of coveted portfolios, the values of ministries are estimated with a lot of uncertainty. Consider, for instance, the Controladoria Geral da União (Comptroller General), whose confidence interval rank extends from 8 through 35, or the Secretaria Especial da Micro e Pequena Empresa (Small Business Administration) whose rank ranges from 4 to 31. This uncertainty is driven both by the number of times an item was evaluated as well as the level of inter-respondent agreement about the portfolio’s desirability. Given this second factor, it is only natural that there is more uncertainty about lower-profile cabinet positions.

These issues can easily be resolved with slightly larger datasets. Consider the rightmost columns in Table 1, which report results from an identical model estimated using data from the larger “expert” survey. The group of ministries that cannot be distinguished from the most valuable one is now reduced to three (from five in the legislator survey), and the average standard error is 0.26 compared to 1.12 in the elite survey. Similarly, the maximum width of the rank-order confidence interval is 9 (compared to 27 in the Congress survey), and the median is now reduced to 6 (compared to 12).

We expected some differences between assessments made by politicians and political scientists, as the former have private information unavailable to even the most informed academics. In the expert poll, Urban Affairs (Cidades) is still the highest ranked ministry among those that we assigned to Tier 2—and in fact is ranked higher than a number of Tier 1 posts—but according to expert opinion it no longer sits in the top cluster of portfolios. Comparison between the two datasets, however, reveals many more similarities than differences. The linear correlation coefficient between the two sets of estimates is ($r = 0.87$) and the estimate of the Spearman
<table>
<thead>
<tr>
<th>Tier</th>
<th>Ministry (Abbreviated Name)</th>
<th>Legislators</th>
<th>Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Worth (α)</td>
<td>(Low–High CI)</td>
<td>Worth (α)</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>1–6</td>
<td>-0.45</td>
</tr>
<tr>
<td>1</td>
<td>-0.15</td>
<td>1–5</td>
<td>-0.22</td>
</tr>
<tr>
<td>1</td>
<td>-0.25</td>
<td>1–5</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>-0.61</td>
<td>1–10</td>
<td>-0.03</td>
</tr>
<tr>
<td>1</td>
<td>-1.23</td>
<td>2–8</td>
<td>-0.37.</td>
</tr>
<tr>
<td>1</td>
<td>-1.40</td>
<td>3–11</td>
<td>-0.38.</td>
</tr>
<tr>
<td>1</td>
<td>-1.67</td>
<td>3–13</td>
<td>-0.01</td>
</tr>
<tr>
<td>1</td>
<td>-2.13*</td>
<td>6–17</td>
<td>-1.09**</td>
</tr>
<tr>
<td>2</td>
<td>-2.17*</td>
<td>5–17</td>
<td>-2.14**</td>
</tr>
<tr>
<td>1</td>
<td>-2.20*</td>
<td>5–18</td>
<td>-0.99**</td>
</tr>
<tr>
<td>1</td>
<td>-2.33*</td>
<td>4–22</td>
<td>-1.44**</td>
</tr>
<tr>
<td>1</td>
<td>-2.41*</td>
<td>6–21</td>
<td>-2.30**</td>
</tr>
<tr>
<td>1</td>
<td>-2.50*</td>
<td>8–21</td>
<td>-1.08**</td>
</tr>
<tr>
<td>1</td>
<td>-2.60*</td>
<td>7–24</td>
<td>-1.35**</td>
</tr>
<tr>
<td>1</td>
<td>-2.75*</td>
<td>7–25</td>
<td>-1.54**</td>
</tr>
<tr>
<td>2</td>
<td>-2.75**</td>
<td>8–22</td>
<td>-0.72**</td>
</tr>
<tr>
<td>2</td>
<td>-2.90*</td>
<td>6–27</td>
<td>-1.47**</td>
</tr>
<tr>
<td>2</td>
<td>-2.95**</td>
<td>10–24</td>
<td>-1.38**</td>
</tr>
<tr>
<td>2</td>
<td>-3.07**</td>
<td>8–26</td>
<td>-0.82**</td>
</tr>
<tr>
<td>4</td>
<td>-3.08*</td>
<td>4–31</td>
<td>-3.01**</td>
</tr>
<tr>
<td>2</td>
<td>-3.19**</td>
<td>12–26</td>
<td>-2.47**</td>
</tr>
<tr>
<td>4</td>
<td>-3.29**</td>
<td>9–28</td>
<td>-2.19**</td>
</tr>
<tr>
<td>1</td>
<td>-3.64**</td>
<td>13–32</td>
<td>-0.68**</td>
</tr>
<tr>
<td>3</td>
<td>-3.70**</td>
<td>17–28</td>
<td>-2.42**</td>
</tr>
<tr>
<td>2</td>
<td>-3.82**</td>
<td>15–32</td>
<td>-2.35**</td>
</tr>
<tr>
<td>3</td>
<td>-3.91**</td>
<td>19–29</td>
<td>-1.76**</td>
</tr>
<tr>
<td>4</td>
<td>-3.94**</td>
<td>11–34</td>
<td>-2.23**</td>
</tr>
<tr>
<td>3</td>
<td>-4.38**</td>
<td>22–33</td>
<td>-2.80**</td>
</tr>
<tr>
<td>4</td>
<td>-4.40**</td>
<td>19–34</td>
<td>-3.39**</td>
</tr>
<tr>
<td>4</td>
<td>-4.65**</td>
<td>26–33</td>
<td>-2.99**</td>
</tr>
<tr>
<td>3</td>
<td>-4.86**</td>
<td>24–36</td>
<td>-2.54**</td>
</tr>
<tr>
<td>3</td>
<td>-5.08**</td>
<td>24–37</td>
<td>-3.05**</td>
</tr>
<tr>
<td>4</td>
<td>-5.13**</td>
<td>26–37</td>
<td>-2.49**</td>
</tr>
<tr>
<td>4</td>
<td>-5.26**</td>
<td>26–37</td>
<td>-4.00**</td>
</tr>
<tr>
<td>3</td>
<td>-5.55**</td>
<td>29–37</td>
<td>-3.22**</td>
</tr>
<tr>
<td>4</td>
<td>-5.61**</td>
<td>28–37</td>
<td>-4.21**</td>
</tr>
<tr>
<td>4</td>
<td>-5.66**</td>
<td>29–37</td>
<td>-3.41**</td>
</tr>
</tbody>
</table>

The BT estimates are presented in column α and ***, *, and . represent p-values that are < 0.01, < 0.05, and < 0.1 for the comparison relative to the baseline (highest worth) item. The last columns report the 90% confidence intervals for the ranks of each ministry, computed by a parametric bootstrap. See text for details.
Figure 1: Comparison of Elite and Expert Opinions

Figure shows a comparison between “elite” and “expert” BT estimates of the worth of parties.

rank correlation is ($\rho = 0.88$). This rank correlation is only slightly weaker if we measure association between the lower- and upper-bounds of the rank confidence intervals ($\rho = 0.82$) and $(0.87)$ respectively). The rank confidence intervals in the two sets of estimates overlap for all portfolios save two.

Interestingly, Small Business is once again the most noisily estimated cabinet-level post, and Comptroller General is not far behind. This suggests that much of the uncertainty around these estimates is driven by real irresolution with regard to the portfolios in question, rather than by any feature of the model or any issue with randomization.

Even though the high valuation of Urban Affairs by legislators surprised us, Figure 1 reinforces our finding of broad similarity between the two surveys. We see that both sets of respondents are in agreement about the most valuable ministries, although legislators discriminate within this subset more strongly than do academics. Both sets of referees also identify ministries of intermediate value (visual inspection shows clusters of medium-high and medium-low units), as well as a common and sizeable set of bottom-feeding portfolios. From a universe of 37, the scatterplot identifies only four real outliers: Foreign Affairs, Small Business, Transportation, and Science and Technology. Only for the latter two portfolios do the rank confidence
intervals in the two datasets not overlap.

4 Robustness to deviations from randomization

One glaring limitation of our elite survey is that because it was partially conducted using pen and paper (the “manual version” of the survey), we could not fully implement the randomization of contests being evaluated. In the manual survey, fewer different contests were presented more frequently than in the electronic version of the survey. There were 221 different contests in the full legislator dataset. In the electronic subset of the data, 179 contests happened only once, 21 twice, and only three appear three or four times. In the manual version, there were a total of only 19 different contests\footnote{Although the four contests in each of the five versions of the paper questionnaire were randomly selected, by pure chance one matchup appeared on two versions.} which were repeated at least 12 times.

This deviation from randomization is less of a problem that it might seem at first. Each additional contest adds little to the estimated worth for two reasons. First, what matters in the model is the probability of victory in a certain contest, and not the number of victories, so increasing the $N$ has limited impact. Moreover, any changes in probabilities that occur with an increase in $N$ are typically in the tails of the distribution, and these tend to matter less because the model implies a logit link function (as defined in Equation 2). Second, recall the way in which the model handles transitivity: the value of “defeating” another item depends on the whole set of interactions between all items. Therefore, additional matchups between the same items have progressively less impact on the estimates of $\alpha$.

To test these intuitions, we simulated the effect of eliminating the excess repeated contests that were generated by the manual survey. We did this by implementing a form of non-parametric bootstrapping in which we sampled only four of each of the 19 high-frequency contests that were included in the manual survey\footnote{We chose four to allow for ties, but results are identical with sampling fewer matchups.}. We appended the drawn contests to all the contests from the online portion of the legislative survey, and estimated the same BT model as before. We repeated this procedure for 1,000 different samples and then took the average point estimate across all simulations to produce a new set of estimates.

The correlation between the worth of portfolios derived from these 1,000 simulations and our original estimates using the legislator survey (reported in Table 1) is a whopping 0.98, which means that the impact of the excess contests is trivial. Moreover, if we do the same simulations drawing fewer than four contests, the correlation coefficients with the original estimates change...
only marginally; they are never lower than 0.94, even if we draw only one occurrence of each matchup. And, if anything, as Figure 2 shows, Urban Affairs would be even more valuable than what we estimated with the full dataset, which implies that the surprise valuation of this portfolio was not driven by the randomization process. Two other significant outliers are Institutional Relations and Science and Technology; the latter, as we observed above, was already known to be one of the noisier estimates in the set. The bottom line here is that the number of repetitions of each contest matters relatively little to the stability of the estimates.

5 Robustness to Subsetting the Data

While respondents to our legislative survey are all seasoned politicians, our pool of political scientists is potentially heterogenous with respect to knowledge of portfolio salience. The expert survey, therefore, is an interesting setting in which to explore the extent to which estimates vary across subsets of respondents.

For this analysis we divided our expert survey into two. In one subset we included 1,647 responses from respondents whose self-declared expertise on studies of the presidency, the leg-
islative branch, or of executive-legislative relations fell above the median for the sample (71 on a 0–100 scale). In the other subset we included 1,454 responses from political scientists with lower subjective competence. We then estimated the same unrestricted BT model on each subset, and compared the retrieved \( \alpha \)'s from each of these subsets with what we had previously estimated from the full sample of experts and from the separate sample of legislators. The estimates produced by high- and low-expertise academics are almost identical to each other \((r = 0.96)\). More importantly, estimates from the two respective subsets correlate very highly with the estimates produced using all expert responses \((0.99\) for each subset) and with those produced by surveying legislators \((0.86\) for each subset). The takeaway message here is that estimates are not particularly sensitive to subsets of the data, and this has operational implications for researchers. Given that it is relatively easier to survey experts than legislators, it might be sensible to opt for the former approach.

6 Estimates vs. Observable Measures

Does our approach to measuring portfolio salience provide us with any information we could not obtain observationally? If we could predict legislators’ or even experts’ assessments solely with objective indicators, then it would not really be necessary to collect primary data for this task. To examine this question, we compiled data on objective characteristics of the ministries that existed in Brazil in 2015. We chose this as the reference date as it was the last moment at which all 37 of our portfolios coexisted.

<table>
<thead>
<tr>
<th>Table 2: Linear Association Between BT Estimates and Observable Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elite BT Estimates</td>
</tr>
<tr>
<td>Log Total Budget</td>
</tr>
<tr>
<td>( P)-value</td>
</tr>
<tr>
<td>InvestBudget</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Appointees</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Policy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Predicted Worth</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

We obtained each ministry’s effective total and investment budget\(^{15}\) as well as the total

\(^{15}\)By “effective” we mean the budget that was appropriated and liquidated, which in Brazil is sometimes significantly different from what was earmarked.
number of political appointments available within each ministry\textsuperscript{16} and the total number of legislative bills authored or co-signed by each the ministry in 2015. Table 2 shows the linear correlation coefficients between the expert and elite based BT estimates (see Table 1) and each of these objective indicators. The correlations are strongest when it comes to political appointees, suggesting that an important dimension of portfolio salience is the value of the ministry as a source of patronage. The correlations are slightly weaker for our measure of policy relevance (involvement in legislation) and, surprisingly, much weaker for our three indicators of budgetary impact.

In fact, even if we combine all indicators into a linear predictor by regressing the BT estimates on all of the observed characteristics, the linear correlation between the predicted values and our estimates does not rise past 0.6. The $R^2$ of these regressions is almost identical for the sets of estimates, and is just under 0.4. This suggests that both experts and politicians have information beyond what objective indicators can reveal. It also suggests that, perhaps, the relative worth of alternative portfolios depends on different observable characteristics that vary (qualitatively) across units. Therefore, even if we correctly identify a menu of characteristics that matter, they are unlikely to matter equally for all portfolios. Deductively defining a model to account for this heterogeneity is probably much harder than simply collecting perceptual data directly from politicians and experts. We suspect that perceptual data already incorporate this elusive cross-unit heterogeneity, and this folk wisdom then informs a series of simple binary choices made in the surveys.

\section{Conclusion}

We propose a portfolio salience measure estimated by applying the Bradley-Terry model to pairwise comparisons of ministries undertaken by relevant informants (either experts or political practitioners). The method’s main advantage is that the relevant data can be collected through simple, intuitive surveys. We show that the way in which legislators rank cabinet posts deviates from what would be predicted by observable characteristics of ministries. This implies that direct ratings are necessary to assess portfolio salience, and these are most easily achieved via a series of head-to-head matchups between concrete alternatives.

Our ongoing research suggests that professional politicians seem to have “private informa-
\textsuperscript{16}These political appointees are known as DAS positions, levels 1 through 6, and are selected at the discretion of ministers. One third of DAS positions must be filled by existing civil servants. Regular civil service positions, in contrast, are filled only by competitive public examination.
tion” about the relative worth of ministries. We also show that experts reach very similar conclusions as legislators, which means that expert surveys including pairwise comparisons of ministries—which tend to be easier to field than surveys with elites—might generate estimates that are “good enough” for most research purposes. Regardless of the pool of informants, the “wisdom of the crowds” seems to trump observation of objective indicators.

References


URL: [https://www.jstatsoft.org/v048/i09](https://www.jstatsoft.org/v048/i09)


Appendix: The Bradley Terry Model

What follows is the textbook derivation of the Bradley-Terry model, available from multiple sources. We start by defining $\pi_{ij}$ as the probability that item $i$ is chosen over item $j$ in a pairwise comparison:

$$\pi_{ij} = \frac{\alpha_i}{\alpha_i + \alpha_j}$$  \hspace{1cm} (3)

We then parametrize the probability as function of $\exp(\alpha)$. Other choices of parametrization are possible, and have been employed over the years, but this original Bradley Terry formulation leads to:

$$\pi_{ij} = \frac{\exp(\alpha_i)}{\exp(\alpha_i) + \exp(\alpha_j)}$$  \hspace{1cm} (4)

We then express $\pi_{ij}$ a log-odds, with a logit transformation

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = \log(\pi_{ij}) - \log(\pi_{ji})$$  \hspace{1cm} (5)

Finally, replacing $\text{logit}(\pi_{ij})$ and with minimal algebra, we arrive at:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\exp(\alpha_i)}{\exp(\alpha_i) + \exp(\alpha_j)}\right) - \log\left(\frac{\exp(\alpha_j)}{\exp(\alpha_i) + \exp(\alpha_j)}\right) = \log(\exp(\alpha_i)) - \log(\exp(\alpha_j))$$  \hspace{1cm} (6)

If we then define $\lambda_i = \exp(\alpha_i)$, we get

$$\text{logit}(\pi_{ij}) = \lambda_i - \lambda_j$$  \hspace{1cm} (7)

Which is the logit formulation of the problem as a function of the exponential transformation of the underlying worth parameters.