Modeling Dynamic Comparative Public Opinion [DRAFT]*

Frederick Solt   *University of Iowa*

The study of public opinion in comparative context has been hampered by data that is sparse, that is, unavailable for many countries and years; incomparable, i.e., ostensibly addressing the same issue but generated by different survey items; or, most often, both. Questions of representation and of policy feedback on public opinion, for example, cannot be explored fully from a cross-national perspective without comparable time-series data for many countries that span their respective times of policy adoption. Recent works (Claassen 2019; Caughey, O'Grady, and Warshaw 2019) have introduced a latent variable approach to the study of comparative public opinion that maximizes the information gleaned from available surveys to overcome issues of sparse and incomparable data and allow comparativists to examine the dynamics of public opinion. This paper advances this field of research by presenting a new model and software for estimating latent variables of public opinion from cross-national survey data that yield superior fit and more quantities of theoretical interest than previous works allow.

Keywords: public opinion, item response theory, measurement

A perennial problem confronting scholars of comparative public opinion is that the exact question relevant to their research topic is only infrequently asked, making response data scarce across countries and over time. Although a wealth of surveys may provide information on the state of public opinion on the issue of interest in different countries over the years, significant hurdles exist to putting all of this information to use in any comparative study, beginning with the incomparability of the multitude of different questions asked.

As a result, much work in the study of comparative public opinion examines one cross-section of countries, typically provided by a single cross-national survey (see, e.g., Ansell 2014; Clayton et al. 2019). Some studies capture some element of change over time by taking advantage of multiple waves of an ongoing cross-national survey (see, e.g., Abou-Chadi and Finnigan 2019; Busemeyer, Abrassart and Nezi 2020) or, less commonly, drawing on multiple surveys that employed the same items (see, e.g., Ezrow and Hellwig 2014; Dalton 2017). Only for a very few topics are scholars able to capitalize on a relevant item that has been frequently asked by the same survey in many countries: see, for example, how Norris (2011, 70-77) took advantage of the Eurobarometer's longstanding

---

questions on trust in government, and Hagemann, Hobolt and Wratil (2017) that program's equally longstanding item on whether European integration is a "good thing."

A growing body of work takes a different tack, examining the dynamics of public opinion in one or a few countries over time. One approach of this sort compiles responses to the same survey questions asked on an annual or near-annual basis (see, e.g., Hobolt and Klemmensen 2008; Soroka and Wlezien 2010). But even narrowing one's focus to only a single country typically gains few items asked with such regularity, and concerns that any single question can adequately capture opinion are well placed. These factors have led to research applying the dyad-ratios algorithm elaborated in Stimson (1991), tremendously influential in the study of American politics, beyond the U.S. case: among others, Bellucci and Pellegata (2017) on Italian policy mood and Jennings et al. (2017) on British political discontent.

In recent years, there have even been applications of Stimson's dyad-ratios method to each of many countries. Carlin et al. (2018), for example, generate time series of presidential approval in eighteen Latin American countries, and Guinaudeau and Schnatterer (2019) generate annual measures of attitudes toward European integration in the countries of the European Union. A shortcoming of this approach, however, is that because each country's series is estimated separately, the results are not assured to be fully comparable across countries.

Latent variable models estimated with Bayesian methods provide a means of overcoming this problem, and they have become a very useful tool for combining the information available in disparate sources into cross-nationally comparable time series. Such models have been used to measure many difficult concepts in comparative politics, from democracy (Treier and Jackman 2008; Pemstein, Meserve and Melton 2010) through governance (Arel-Bundock and Mebane 2011) and state capacity (Hanson and Sigman 2019) to judicial independence (Linzer and Staton 2015) and respect for human rights (Fariss 2014). Two recent works, Claassen (2019) and Caughey, O'Grady and Warshaw (2019), have introduced a Bayesian latent variable approach to the study of comparative public opinion that maximizes the information gleaned from available surveys to overcome issues of sparse and incomparable data and allow comparativists to examine the dynamics of public opinion.

This paper presents DCPO, a new model and software for estimating Bayesian latent variables of comparative dynamic public opinion from cross-national survey data. The DCPO model

yields superior fit and more quantities of theoretical interest than the models offered in previous works allow; the `DCPO` package for R estimates this model. The accompanying package `DCPOtools` facilitates the formidable data management tasks involved: collecting and managing the survey datasets, identifying the country-years in which respondents were contacted in each, and preparing the response data for the model.

The paper begins with a detailed description of the DCPO model, building up from the individual level. It continues with a comparison of the performance of the DCPO model to the alternate models of public opinion presented in Claassen (2019) and Caughey, O'Grady and Warshaw (2019). The paper then describes the workflow involved in generating estimates of dynamic comparative public opinion with the `DCPO` software and employing them in analyses.

Estimating Dynamic Comparative Public Opinion

The logic underlying DCPO's approach to estimating dynamic comparative public opinion starts at the individual level with the two-parameter logistic item response theory ("2PL IRT") model. In this model, the probability that individual i responds affirmatively to a dichotomous question q is a function of the individual's score on the unbounded latent trait, θ'_i, plus two parameters that characterize the question, its *difficulty*, β_q, and its *dispersion*, α_q:

$$\Pr(y_{iq} = 1) = \text{logit}^{-1}(\frac{\theta'_i - \beta_q}{\alpha_q}) \tag{1}$$

Figure 1 shows how these parameters interact using simulated data. First, both panels show that regardless of the question, as an individual's score on the unbounded latent trait, θ'_i, increases, the individual becomes more likely to give an affirmative response. Second, the left panel describes how the probability of answering affirmatively changes with question difficulty, β_q: some questions demand more of respondents than others. Take, for example, questions regarding abortion. One need not hold views very favorable toward abortion rights to agree that abortion should be permitted when the life of the woman is at risk. The question is not very difficult, and only those who view abortion very unfavorably will answer negatively. To respond affirmatively to a question that asks whether abortion should be permitted for a single woman who does not want to marry the man requires respondents to have more favorable views toward abortion; this question is more difficult.

3

The question of whether one *approves* of abortion in this same situation is still more difficult, that is, some people who accept that abortion should be permitted in such cases will nevertheless express disapproval of women who exercise their right. The left panel illustrates this graphically: it shows that individuals with the same score on the latent trait are less likely to respond affirmatively to a question as the question's difficulty increases (here, all three questions depicted have a dispersion of 1). When θ'_i and β_q are equal, the probability of an affirmative response is 50%.
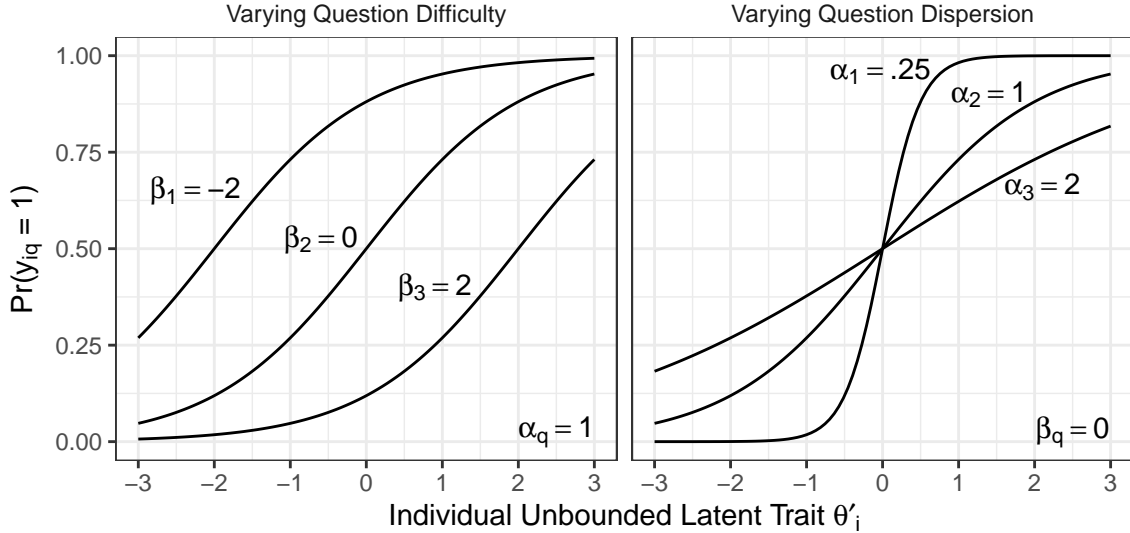
Figure 1: Probability of an Affirmative Response in the 2PL IRT Model

Third, the right panel of Figure 1 portrays how the *dispersion* of the question, α_q, affects the probability of positive responses. A question's dispersion is proportional to its measurement error with regard to the latent variable; it may be larger because the question is confusing to some respondents, or because the question does not cleanly map onto the latent trait. For example, a question that asks respondents about laws protecting a population from employment discrimination may tap views about workplace protections as much or more than it does attitudes toward the particular group. If it is nevertheless used as an indicator of the latter, this question will have a high dispersion. As shown in the right panel, as question dispersion increases, individuals with lower scores on the latent trait become more likely to respond affirmatively and those with higher scores on the latent trait are more likely to respond negatively; in other words, the slope of the curve describing the relationship between the latent trait and the probability of answering affirmatively flattens out (here, all three questions depicted have a difficulty of 0).

4

To aggregate individual-level responses to generate a population-level estimate of public opinion is a matter of integration. If scores on the unbounded latent trait are modeled as normally distributed within the population of country k at time t, integrating Equation 1 yields the population-level two-parameter logistic IRT model:

$$\eta_{ktq} = \text{logit}^{-1}\left(\frac{\bar{\theta}'_{kt} - \beta_q}{\sqrt{\alpha_q^2 + (1.7 * \sigma_{kt})^2}}\right) \tag{2}$$

where η_{ktq} is the expected probability that a random person in country k at time t answers question q affirmatively and $\bar{\theta}'_{kt}$ and σ_{kt} are the mean and standard deviation of the unbounded latent trait θ' in the population of country k at time t (see Mislevy 1983, 280; see also, using the two-parameter probit IRT model, McGann 2014, 120).

Further, many survey questions are not simply dichotomous, but instead measure respondents' attitudes on ordinal scales, assumed here to range from 1 to R. To take advantage of this additional information, DCPO uses the cumulative logit formulation, in which the probability to be estimated is not that individual i provided a simple affirmative response, as in Equation 2 above, but instead the probability that individual i provided a response at least as positive as response r for all r greater than 1 and less than or equal to R. This yields a population-level graded response model:

$$\eta_{ktqr} = \text{logit}^{-1}\left(\frac{\bar{\theta}'_{kt} - \beta_{qr}}{\sqrt{\alpha_q^2 + (1.7 * \sigma_{kt})^2}}\right) \tag{3}$$

Again, the differences between Equation 3 for the population-level graded response model and Equation 2 for the population-level two-parameter logistic IRT model above are in the additional subscripts for r to η and β. In Equation 3, η_{ktqr} is the expected probability that a random individual in country k at time t replies to question q with a response at least as positive as response r. And the additional subscript to β_{qr} indicates that this parameter represents the difficulty of response r of question q, constrained to be increasing for increasing r. That is to say, for example, that the model makes the straightforward assumption that the response of "strongly agree" to a question will, on average across respondents, reflect higher levels of the latent trait than the response of "agree" to that same question.

A final addition to the DCPO model of the probability η_{ktqr} takes into account differences

in item response bias across countries. Responses to survey questions may vary across different countries not only as a result of differences in attitudes and preferences but also due to translation issues (see, e.g., Davidov and De Beuckelaer 2010), cultural differences in acquiescence and extreme response styles (see, e.g., van Herk, Poortinga and Verhallen 2004), or other idiosyncrasies—recall Tarrow's (1971, 344) famous observation that the French understood survey questions regarding their 'interest in politics' as inquiring about the strength of their partisan affiliations. Rather than simply allowing such problems of equivalence to contribute to the error of the model, they can be addressed in part by explicitly modeling the country-specific item bias (Stegmueller 2011). In the context of the population-level graded response model presented in Equation 3, this can be done by including country-specific item bias as a country-varying shift in the difficulty of each question, denoted here as δ_{kq}:

$$\eta_{ktqr} = \text{logit}^{-1}(\frac{\bar{\theta}'_{kt} - (\beta_{qr} + \delta_{kq})}{\sqrt{\alpha_q^2 + (1.7 * \sigma_{kt})^2}})\tag{4}$$

Estimating δ_{kq} in Equation 4 requires repeated administrations of question q in country k. When responses to question q are observed in country k in only a single year, the DCPO model sets δ_{kq} to zero by assumption. This means that incorporating these 'one-shot' surveys comes at the price of increasing the error of the model by any country-item bias that is present.[1] Of course, questions that are asked repeatedly over time in only a single country pose no risk of country-specific item bias, so δ_{kq} in such cases can be set to zero without harm.

Before continuing, it is useful to pause to briefly review the relationships of the parameters in this equation. The numerator of Equation 4 implies first that η_{ktqr}, the expected probability of a response at least as positive as r, increases as the mean value of the unbounded latent trait in the population, $\bar{\theta}'_{kt}$, increases and second that η_{ktqr} decreases with more difficult questions and higher response categories, β_{qr}, and particularly so where the item-response bias, δ_{kq}, is positive. The equation's denominator implies that η_{ktqr} is drawn closer to 50% as the geometric mean of the question's dispersion, α_q, and the standard deviation in opinion in the population, σ_{kt}, increases: individuals sampled even from a population with a negative mean value on the unbounded latent trait and answering a difficult question will still be more likely to supply more positive responses—

---

[1]Researchers uncomfortable with this price may follow Claassen's (2019, 5-6) approach with regard to country-specific item bias, which is to instead discard survey data on questions that were not administered more than once in a given country at the cost of giving up all of the information contained in these data to guard against the mere risk of introducing some additional error to the estimates.

and those from a population with a positive mean score on the unbounded latent trait and answering an easy question will be more likely to supply more negative responses—if the question's dispersion is greater or attitudes are more polarized in the population.

Given this expected probability η_{ktqr}, the total number of survey responses at least as positive as r to each question q in country k at time t, y_{ktqr}, out of the total number of respondents surveyed, n_{ktqr}, is then modeled using the beta-binomial distribution, which allows for an overall dispersion parameter, ϕ, to account for additional sources of survey error (see McGann 2014, 120; Claassen 2019, 4-5).

$$a_{ktqr} = \phi \eta_{ktqr} \tag{5}$$

$$b_{ktqr} = \phi(1 - \eta_{ktqr}) \tag{6}$$

$$y_{ktqr} \sim \text{BetaBinomial}(n_{ktqr}, a_{ktqr}, b_{ktqr}) \tag{7}$$

Then, to account for the dynamics of comparative public opinion—the change over time—the prior distributions for the public opinion parameters in Equation 4, $\bar{\theta}'_{kt}$ and σ_{kt}, are given by simple local-level dynamic linear models. For $\bar{\theta}'_{kt}$, this random-walk prior takes the following form:

$$\bar{\theta}'_{kt} \sim \text{N}(\bar{\theta}'_{k,t-1}, \sigma^2_{\bar{\theta}'}) \tag{8}$$

By treating these parameters' values at time $t - 1$ as their expected values at time t, this prior works to smooth estimates of the mean of the latent trait in the population of each country over time. For σ_{kt}, a lognormally distributed prior is used:

$$\sigma_{kt} \sim \text{LN}(\log(\sigma_{k,t-1}), \sigma^2_\sigma) \tag{9}$$

The lognormal distribution ensures that the prior for σ_{kt} takes on only positive values. In both cases, if no survey data is available for a particular time, these dynamic models provide estimates based on the estimates for previous and subsequent periods. The variances $\sigma^2_{\bar{\theta}'}$ and σ^2_σ are estimated from the data.

Previous efforts to measure cross-national aggregate public opinion as a latent variable have generated estimates on an unbounded scale with mean zero and unit variance (see Claassen 2019,

14; Caughey, O'Grady and Warshaw 2019, 8), and this is the scale of $\bar{\theta}'_{kt}$. However, like many other concepts in political science (see Linzer and Staton 2015, 229), many aspects of public opinion are conceptually bounded, that is, it make sense to think of them as lying along a scale from fully absent to fully present in the relevant public. Take attitudes toward immigration as an example. One hypothetical country's citizens are, at a given time, absolutely opposed to any immigration, while another's are completely welcoming to migrants. Actual countries' levels of public opinion toward immigration are better understood not as unbounded but as falling somewhere between the bounds described by these two hypothetical countries. It is not surprising, then, that earlier works aggregating public opinion within a single country presented results on bounded scales (see Stimson 1991; McGann 2014). Even if the issue in question is less easily viewed as bounded, bounding is still a good idea because it reduces the uncertainty for the estimates for countries at the extremes, that is, those countries whose values should in fact be easier to estimate. As Linzer and Staton (2015, 229) note, "bounding the latent variable may do little harm to the scale and produce more sensible estimates of uncertainty." DCPO therefore uses the logistic function to transform the unbounded estimates, $\bar{\theta}'_{kt}$, to the unit interval:

$$\bar{\theta}_{kt} = \text{logit}^{-1}(\bar{\theta}'_{kt} - 1) \tag{10}$$

The resulting $\bar{\theta}_{kt}$ is the DCPO estimate of the mean public opinion for country k at time t. Together, σ_{kt}, which provides a measure of polarization in public opinion, and $\bar{\theta}_{kt}$ will typically be the main quantities of interest.

The model is identified by imposing a series of constraints that fix location, direction, and scale. The dispersion parameters α are constrained to be positive—and all survey responses are coded to have the same polarity—to fix direction. One researcher-specified difficulty parameter β is set to a value of 1 to identify location, and, as mentioned previously, for each question q the difficulties for increasing response values r are constrained to be increasing. To ensure identification, the sum of δ_{kq} across all countries k is fixed to zero for each question q:

$$\sum_{k=1}^{K} \delta_{kq} = 0 \tag{11}$$

Weakly informative priors are placed on most parameters. The dispersion parameters α_q are drawn from standard half-normal prior distributions, that is, the positive half of N(0, 1). The first difficulty parameters for each question, β_{q1}, are drawn from standard normal prior distributions, and the differences between βs for each r for the same question q are drawn from standard half-normal prior distributions. The item-bias parameters δ_{kq} receive normally-distributed hierarchical priors with mean 0 and standard deviations drawn from standard half-normal prior distributions. The initial value of the mean unbounded latent trait for each country, $\bar{\theta}'_{k1}$, is assigned a standard normal prior, as are the transition variances $\sigma^2_{\bar{\theta}'}$ and σ^2_σ; the initial value of the standard deviation of the unbounded latent trait for each country, σ_{k1}, is drawn from a standard lognormal prior distribution. The overall dispersion, ϕ, receives a somewhat more informative prior drawn from a gamma(4, 0.1) distribution that yields values that are well scaled for that parameter. The model is estimated using the `DCPO` package for R, which is written in the Stan probabilistic programming language (Stan Development Team 2019a,b).

Table 1: Comparing Latent Variable Models of Public Opinion

	McGann (2014)	Claassen (2019) Model 5	Caughey, O'Grady, and Warshaw (2019)	DCPO
Cross-National	No	Yes	Yes	**Yes**
Dynamic Priors	No	Yes	Yes	**Yes**
Ordinal	No	No	Yes	**Yes**
Country-Varying Question Difficulty	No	Yes	No	**Yes**
Bounded Mean Public Opinion	Yes	No	No	**Yes**
Country-Year Public Opinion Std. Deviation	Yes	No	No	**Yes**

Table 1 summarizes the DCPO model just described in comparison to three alternate ways of modeling public opinion: that presented by McGann (2014), Claassen's (2019) preferred Model 5, and that employed in Caughey, O'Grady and Warshaw (2019). It reveals that DCPO incorporates the desirable features from each of those three models. DCPO generates estimates of public opinion that are comparable not only over time but also across countries, takes advantage of all of the information encoded in responses to ordinal questions, explicitly accounts for country-item bias by

including terms to capture country-varying question difficulty, provides easily interpreted bounded estimates of mean public opinion, and includes the standard deviation of public opinion by country-year to represent the extent of polarization. The next section provides a demonstration of how these features allow DCPO to provide a better fit to actual data.

Assessing Relative Fit

To assess the ability of the DCPO model to fit public opinion data relative to the two alternative approaches, I use the set of survey questions on support for democracy employed in Claassen (2019, 7-8).

The first three columns of Table 2 present the results of an internal validation test, that is, a test that uses the same data that was used to fit the model (see, e.g., Claassen 2019, 9). In column 1, the mean absolute error (MAE) measures the average difference between the observed proportion of survey respondents in country k in year t with replies to question q with a response at least as positive as response r (or with an affirmative response, in the case of Claassen's (2019) Model 5) and the model's predicted proportion across all countries, years, questions, and response categories. Given, however, that Claassen's (2019) Model 5 is fit to dichotomized data, while the Caughey, O'Grady and Warshaw (2019) and DCPO models are fit to the original, possibly ordinal, survey data, which have higher variance, comparing the MAE across all three models can be somewhat misleading. Therefore, the MAE of the country means—the average proportions answering affirmatively (for Claassen's (2019) Model 5) or with a response at least as positive as response r (for Caughey, O'Grady and Warshaw (2019) and DCPO) in each country across all years and all questions—serve as a baseline (see Claassen 2019, 11-12); these appear in column 2. The percentage reduction in MAE achieved by each model, listed in column 3, represents the improvement in fit compared to this baseline.

The last three columns of Table 2 present the results of an external validation test, a test of the models' ability to predict out-of-sample survey responses. This test employs k-fold validation with 10 folds, that is, it randomly divides the available country-year-questions into tenths and then sequentially treats each tenth as a test set to be predicted while fitting the model on a training set consisting of the other nine-tenths of the data. Column 4 shows the mean of the MAEs of the ten

10

resulting sets of out-of-sample predictions, while column 5 presents the mean improvement over of these ten MAEs over their respective ten country-mean MAEs. Column 6 reports the discrepancy between the percentage of all response proportions that fall within their corresponding predictions' 80% confidence intervals and the expected 80%; it therefore provides a gauge of the accuracy of the models' estimates of uncertainty. A negative value in this column indicate that less than 80% of the observed out-of-sample survey proportions are included with the model's predictions' nominal 80% credible intervals, and so its uncertainty estimates are overconfident, while a positive value indicates the opposite, that more than 80% of the observed proportions fall within the credible intervals and so the model's uncertainty estimates are overly conservative.

Table 2: Internal and External Validation Tests

	Internal Validation Test			External Validation Test		
	(1) Mean Absolute Error	(2) Country-Means	(3) % Improvement in	(4) k-fold Mean	(5) k-fold Mean % Improve-	(6) k-fold 80% Credible Interval
Model	(MAE)	MAE	MAE	MAE	ment	Coverage
Claassen (2019) Model 5	0.032	0.112	71.4	0.057	51.7	+4.9
Caughey, O'Grady, and Warshaw (2019)	0.049	0.186	73.7	0.063	66.1	-67.4
DCPO	**0.031**	**0.186**	**83.3**	**0.055**	**70.5**	**-4.5**

The internal validation test uses the same data for model fitting and validation; the external validation test employs k-fold validation with 10 folds, randomly dividing the data into tenths and then sequentially treating each tenth as a test set while fitting the model on a training set consisting of the remaining 90 percent of the data. Percent improvement in MAE (column 3) compares the model's MAE (column 1) and the corresponding country-means MAE (column 2).

Comparing first the results for Claassen's (2019) Model 5 with those for the Caughey, O'Grady and Warshaw (2019) model reveals that, at least for this set of survey questions, while the former has a smaller MAE, the latter accounts for a larger percentage of the variation in its source data left unexplained by its respective country-means model. This is true both in the internal validation test and on average across the ten folds of the external validation test. Again, this discrepancy is possible due to the greater variance in the survey data when its ordinal nature is preserved, as in the Caughey, O'Grady and Warshaw (2019) model, rather than dichotomized, as in Claassen's (2019) approach. An examination of column 6 reveals that the Caughey, O'Grady and Warshaw

(2019) model yielded predictions with credible intervals that are much too narrow, encompassing only about 13% of the actual sample observations, while those for Claassen's (2019) Model 5 were slightly too conservative.[2]

However, it is the DCPO model that provides what is unambiguously the best fit. It features the smallest mean absolute error: on average, in the internal validation test its predictions are just over 3 percentage points away from the actual sample observations, and in the external validation test they are only 5.5 percentage points away from the actual out-of-sample observations. In both cases, these represent the largest percentage improvement over the country-means MAE. Further, in the external validation test, the DCPO credible intervals come closest to matching the nominal 80% level of any of the three models.

Preparing and Using DCPO Estimates of Public Opinion

The old saw of data science—that at least 80% of any project consists of wrangling the data (see, e.g., Lohr 2014)—applies with full force when conducting research on public opinion across countries and over time. Fortunately, the `DCPOtools` package facilitates the formidable data management tasks involved: collecting and managing the survey datasets, identifying the country-years in which respondents were contacted in each, and preparing the response data for the model.

To generate estimates using `DCPO`, one must first identify a list of questions that tap the desired aspect of public opinion, the order of their respective response categories, and the surveys in which they were asked. This is an irreducibly labor-intensive task. But subsequent data managment chores are largely automated.

`DCPOtools` contains information on the country-years in which respondents were surveyed for over 600 survey datasets—users can easily add more—and it takes advantage of the work of Arel-Bundock, Enevoldsen and Yetman (2018) to standardize the names of countries across datasets. To facilitate collaboration and replicability, the package automates, with a single command, the task of downloading needed survey datasets from the ICPSR Data Archive (Solt and Gracey 2016), the Pew Research Center (Solt and Hu 2016), the Dataverse Network (Leeper 2017), the Roper Center for Public Opinion Research (Solt 2017), the European Social Survey (Cimentada 2019), the GESIS

---

[2]The relatively poor uncertainty estimates of the Caughey, O'Grady and Warshaw (2019) model are similar to those found for the dichotomous version presented in Caughey and Warshaw (2015) by Claassen (2019, 12).

Data Archive (Solt 2020a), and many other sources. To be sure, there are many surveys for which downloads cannot be automated—for example the Asian Barometer, which is available only via email—but this step nevertheless greatly speeds the work of dataset acquisition. An additional line of code generates the raw dataset: the weighted number of respondents in each country-year who provided each response to each question in each survey in the researcher's list. One last function reformats the data as input to estimate the DCPO model with the `DCPO` package.

After estimating the DCPO model, researchers should be aware that while such latent variable models seek to make the most of the data available, they cannot completely eradicate the issues of data scarcity and incomparability. Rather, where and when data are more sparse, the credible intervals of the model estimates will—and should—expand to reflect our greater uncertainty as to the state of public opinion. When using the resulting public opinion estimates in analyses, this uncertainty needs to be taken into account. This means not simply employing the point estimates, that is, the means of the posterior distributions (*contra* Claassen 2020a,b).[3]

Instead, researchers should follow the recommendations of work using other latent variables (e.g., Schnakenberg and Fariss 2014; Crabtree and Fariss 2015) and other measures incorporating uncertainty, such as the Standardized World Income Inequality Database (Solt 2020b): generate many duplicate versions of the analysis dataset, assign to each a different random draw from the posterior distributions of the variables measured with uncertainty, perform the analyses repeatedly on each of these multiple versions of the dataset, and combine the results following the rules set out in Rubin (1987). The functional programming tools in the `purrr` package (Henry and Wickham 2019) make this a matter of just a few additional lines of code. Although it is intuitive that neglecting measurement error can lead to false positive results, the example presented by Crabtree and Fariss (2015) demonstrates that doing so can lead to false negative results as well. Quantifying the uncertainty in the measures is one of the strengths of this sort of latent variable estimation, and researchers should be sure to take advantage of it.

---

[3]The analysis in O'Grady and Abou-Chadi (2019) appears to do the same, but the replication materials for that article are as yet not accessible to confirm.

13

Conclusions

The development of latent variable models that identify how public opinion varies both across countries and over time has the potential to trigger a new wave of research on the causes and consequences of public opinion that will take into account the experiences of many countries. This will be crucial for developing and testing our understanding of, for example, not merely if policy change leads to changes in public opinion, but to the conditions in which accelerating, self-reinforcing, or self-undermining feedback (in the useful typology proposed by Busemeyer, Abrassart and Nezi 2020) are more likely to occur.

By examining a broad sample of democracies, DCPO helps researchers avoid conclusions based on the idiosyncrasies of any given political setting and provide a firmer grounding for our understanding of how democracies work and the threats to representation that they face.

References

Abou-Chadi, Tarik and Ryan Finnigan. 2019. "Rights for Same-Sex Couples and Public Attitudes Toward Gays and Lesbians in Europe." *Comparative Political Studies* 52(6):868–895.

Ansell, Ben. 2014. "The Political Economy of Ownership: Housing Markets and the Welfare State." *American Political Science Review* 108(2):383–402.

Arel-Bundock, Vincent, Nils Enevoldsen and CJ Yetman. 2018. "countrycode: Convert Country Names and Country Codes." *Journal of Open Source Software* 3(28):848–849.

Arel-Bundock, Vincent and Walter R. Mebane. 2011. "Measurement Error, Missing Values and Latent Structure in Governance Indicators." Presented at the Annual Meeting of the American Political Science Association.

Bellucci, Paolo and Alessandro Pellegata. 2017. "Citizens' Policy Mood, Policies and Election Outcomes in Italy." *Contemporary Italian Politics* 9(1):8–29.

Busemeyer, Marius R, Aurélien Abrassart and Roula Nezi. 2020. "Beyond Positive and Negative: New Perspectives on Feedback Effects in Public Opinion on the Welfare State." *British Journal of Political Science* .

Carlin, Ryan E., Jonathan Hartlyn, Timothy Hellwig, Gregory J. Love, Cecilia Martínez-Gallardo and Matthew M. Singer. 2018. "Public Support for Latin American Presidents: The Cyclical Model in Comparative Perspective." *Research & Politics* 5(3):1–8.

Caughey, Devin and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model." *Political Analysis* 23(2):197–211.

Caughey, Devin, Tom O'Grady and Christopher Warshaw. 2019. "Policy Ideology in European Mass Publics, 1981–2016." *American Political Science Review* 113(3):674–693.

Cimentada, Jorge. 2019. "essurvey: Download Data from the European Social Survey on the Fly." Available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=essurvey.

Claassen, Christopher. 2019. "Estimating Smooth Country–Year Panels of Public Opinion." *Political Analysis* 27(1):1–20.

Claassen, Christopher. 2020*a*. "Does Public Support Help Democracy Survive?" *American Journal of Political Science* .

Claassen, Christopher. 2020*b*. "In the Mood for Democracy? Democratic Support as Thermostatic Opinion." *American Political Science Review* 114(1):36–53.

Clayton, Amanda, Cecilia Josefsson, Robert Mattes and Shaheen Mozaffar. 2019. "In Whose Interest? Gender and Mass–Elite Priority Congruence in Sub-Saharan Africa." *Comparative Political Studies* 52(1):69–101.

Crabtree, Charles D. and Christopher J. Fariss. 2015. "Uncovering Patterns Among Latent Variables: Human Rights and De Facto Judicial Independence." *Research & Politics* 2(3):1–9.

Dalton, Russell J. 2017. *The Participation Gap: Social Status and Political Inequality.* Oxford: Oxford University Press.

Davidov, Eldad and Alain De Beuckelaer. 2010. "How Harmful are Survey Translations? A Test with Schwartz's Human Values Instrument." *International Journal of Public Opinion Research* 22(4):485–510.

Ezrow, Lawrence and Timothy Hellwig. 2014. "Responding to Voters or Responding to Markets? Political Parties and Public Opinion in an Era of Globalization." *International Studies Quarterly* 58(4):816–827.

Fariss, Christopher J. 2014. "Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297–318.

Guinaudeau, Isabelle and Tinette Schnatterer. 2019. "Measuring Public Support for European Integration Across Time and Countries: The 'European Mood' Indicator." *British Journal of Political Science* 49(3):1187–1197.

Hagemann, Sara, Sara B. Hobolt and Christopher Wratil. 2017. "Government Responsiveness in the European Union: Evidence From Council Voting." *Comparative Political Studies* 50(6):850–876.

Hanson, Jonathan K. and Rachel Sigman. 2019. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." Working paper.

Henry, Lionel and Hadley Wickham. 2019. "purrr: Functional Programming Tools." Available at the Comprehensive R Archive Network (CRAN). `https://CRAN.R-project.org/package=purrr`.

Hobolt, Sara Binzer and Robert Klemmensen. 2008. "Government Responsiveness and Political Competition in Comparative Perspective." *Comparative Political Studies* 41(3):309–337.

Jennings, Will, Nick Clarke, Jonathan Moss and Gerry Stoker. 2017. "The Decline in Diffuse Support for National Politics: The Long View on Political Discontent in Britain." *Public Opinion Quarterly* 81(3):748–758.

Leeper, Thomas J. 2017. "dataverse: R Client for Dataverse 4." Available at the Comprehensive R Archive Network (CRAN). `https://CRAN.R-project.org/package=dataverse`.

Linzer, Drew A. and Jeffrey K. Staton. 2015. "A Global Measure of Judicial Independence, 1948–2012." *Journal of Law and Courts* 3(2):223–256.

Lohr, Steve. 2014. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights." *New York Times* August 14(Section B):4.

McGann, Anthony J. 2014. "Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm." *Political Analysis* 22(1):115–129.

Mislevy, Robert J. 1983. "Item Response Models for Grouped Data." *Journal of Educational Statistics* 8(4):271–288.

Norris, Pippa. 2011. *Democratic Deficit: Critical Citizens Revisited.* New York: Cambridge University Press.

O'Grady, Tom and Tarik Abou-Chadi. 2019. "Not So Responsive After All: European Parties Do Not Respond to Public Opinion Shifts Across Multiple Issue Dimensions." *Research & Politics* 6(4):1–7.

Pemstein, Daniel, Stephen A. Meserve and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.

Schnakenberg, Keith E. and Christopher J. Fariss. 2014. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods* 2(1):1–31.

Solt, Frederick. 2017. "ropercenter: Reproducible Data Retrieval from the Roper Center Data Archive." Available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=ropercenter.

Solt, Frederick. 2020*a*. "gesisdata: Reproducible Data Retrieval from the GESIS Data Archive." Available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=gesisdata.

Solt, Frederick. 2020*b*. "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database." *Social Science Quarterly* 101(3):1183–1199. SWIID Version 8.3, May 2020.

Solt, Frederick and Kellen Gracey. 2016. "icpsrdata: Reproducible Data Retrieval from the ICPSR Archive." Available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=icpsrdata.

Solt, Frederick and Yue Hu. 2016. "pewdata: Reproducible Retrieval of Pew Research Center Datasets." Available at the Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=pewdata.

Soroka, Stuart Neil and Christopher Wlezien. 2010. *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge: Cambridge University Press.

Stan Development Team. 2019*a*. "RStan: the R interface to Stan." R package version 2.19.2. https://CRAN.R-project.org/package=rstan.

Stan Development Team. 2019*b*. "Stan User's Guide, Version 2.21." https://mc-stan.org/docs/2_21/stan-users-guide-2_21.pdf.

Stegmueller, Daniel. 2011. "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* 19(4):471–487.

Stimson, James A. 1991. *Public Opinion in America: Moods, Cycles, and Swings*. Boulder: Westview Press.

Tarrow, Sidney. 1971. "The Urban-Rural Cleavage in Political Involvement: The Case of France." *American Political Science Review* 65(2):341–357.

Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

van Herk, Hester, Ype H. Poortinga and Theo M. M. Verhallen. 2004. "Response Styles in Rating Scales: Evidence of Method Bias in Data From Six EU Countries." *Journal of Cross-Cultural Psychology* 35(3):346–360.