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 missing 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 wealth of surveys provide information on the state of public opinion on various issues in different countries over the years, but scholars have faced significant hurdles to putting all of this information to use in any comparative study. The most challenging of these obstacles is that, across countries and over time, the questions asked regarding any given issue are rarely the same, making responses to these questions incomparable.

As a result, the most common approach to the study of comparative public opinion is to use a single cross-section, typically provided by a single cross-national survey (see, e.g., Dalton, Farrell and McAllister 2011; Ansell 2014). Some works have captured some element of change over time by taking advantage of multiple waves of an ongoing cross-national survey (see, e.g., Inglehart 1997; Inglehart and Welzel 2005) or, more rarely, drawing on multiple surveys that employed the same item (see, e.g., Solt 2011; Ezrow and Hellwig 2014).

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 (or "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:

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

*The paper's revision history and the materials needed to reproduce its analyses can be found on Github here. I am grateful for comments received at the 2014 meetings of the European Political Science Association and American Political Science Association, the 2016 meeting of the Midwest Political Science Association, and at Princeton University, Oxford University, the University of Tennessee, the University of Iowa, Central European University, and the Polish Academy of Sciences. Corresponding author: frederick-solt@uiowa.edu. Current version: November 23, 2019.

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 reveals that individuals with the same score on the latent trait are less likely to respond affirmatively to a question as the question's difficulty, β_q, 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%. Third, the right panel shows that as the *dispersion* of the question, α_q, which is proportional to its measurement error, 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.

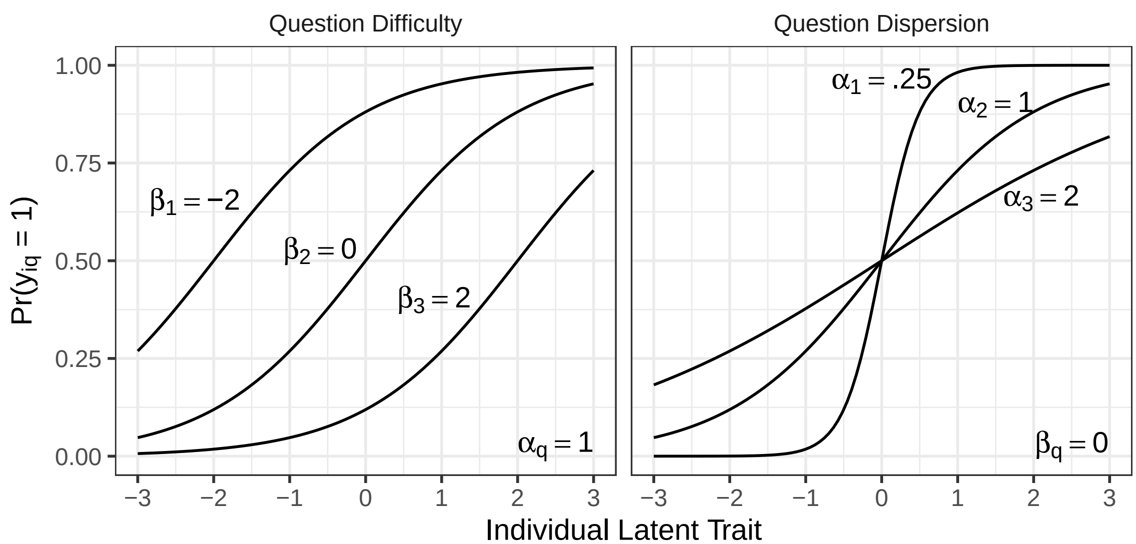

Figure 1: Probability of an Affirmative Response in the 2PL IRT Model

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; Caughey and Warshaw 2015, 200).

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.

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 explicitly by modeling the country-specific item bias (Stegmueller 2011), denoted here as δ_{kq}.[1] In the context of the population-level graded response model presented in Equation 3, including country-specific item bias can be readily understood as a country-varying shift in the difficulty of each question:

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

Before continuing, we 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 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 latent trait and answering a difficult question will still be more likely to supply more positive responses—and those from a population with a positive mean score on the 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).

[1] Claassen's (2019, 5-6) work employs this technique, and because estimating δ_{kq} requires data from repeated administrations of question q in country k, it discards survey data that does not meet this requirement. DCPO also incorporates δ_{kq} to capture country-specific item bias, but it takes a slightly different approach that aims to maximize the incorporation of available survey data: when responses to question q are observed in country k in only a single year, δ_{kq} is set to zero by assumption. This means that incorporating these 'one-shot' surveys will come at the cost of increasing the error of the model by any country-item bias that is present.

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

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

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

Then, to estimate 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:

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

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

By treating these parameters' values at time $t-1$ as their expected values at time t, these priors work to smooth estimates of both the mean and the standard deviation of the latent trait in the population of each country over time. If no survey data is available for a particular time, these 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 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 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 more informative prior drawn from a gamma(4, 0.1) distribution that yields values well-scaled for that parameter.

%McGann2014, 125: "the IRT model has the added advantages of being better justified in terms of individual-level behavior and the additional parameters (the polarization of the population in terms of policy mood) are substantively interesting."

Table 1: Comparing Models of Public Opinion

	McGann (2014)	**Claassen (2019)**	**Caughley, O'Grady, and Warshaw (2019)**	**DCPO**
Cross-National	No	Yes	Yes	**Yes**
Ordinal	No	No	Yes	**Yes**
Country-Varying Question Difficulty	No	Yes	No	**Yes**
Bounded	Yes	No	No	**Yes**
Country-Year Population Variance	Yes	No	No	**Yes**

Conclusions

Single-country studies of public opinion have flourished since the release of Stimson's (1991) algorithm for identifying the common trends in any collection of survey questions that have been repeatedly asked over many years. By extending this work to allow the creation of cross-national pooled time series that identify how public opinion varies both across countries and over time, DCPO 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.

Further, this allows a broadly comparative approach that is new to work on the relationship between opinion and policy. Existing studies on that examine this topic over time investigate only a single country or, much more rarely, a handful of countries. 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.

Table 2: Internal and External Validation Tests

	Internal Validation Tests			External Validation Tests		
Model	**Mean Absolute Error (MAE)**	**Country-Means MAE**	**% Improvement in MAE**	**k-fold Mean MAE**	**k-fold Mean % Improvement**	**k-fold 80% Credible Interval Coverage**
Claassen (2019)	0.032	0.110	70.9	0.056	32.6	85.6
Caughey, O'Grady, and Warshaw (2019)	0.046	0.186	75.3			
DCPO	**0.031**	**0.186**	**83.3**	**0.055**	**69.1**	**75.2**

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 compares the model's MAE (column 1) and the corresponding country-mean MAE (column 2).

References

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

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* pp. 1–20.

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

Dalton, Russell J., David M. Farrell and Ian McAllister. 2011. *Political Parties and Democratic Linkage: How Parties Organize Democracy.* 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.

Inglehart, Ronald. 1997. *Culture Shift in Advanced Industrial Society.* Princeton: Princeton University Press.

Inglehart, Ronald and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy.* Cambridge: Cambridge University Press.

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.

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.

Solt, Frederick. 2011. "Diversionary Nationalism: Economic Inequality and the Formation of National Pride." *Journal of Politics* 73(3):821–830.

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.

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.