# Supplementary Materials

## Using Joint Scaling Methods to Study Ideology and Representation: Evidence from Latin America

March 12, 2015

1. Survey Information (bridge questions and response rates)
2. Selection-corrected estimates
3. Probability of voting for the government's party
4. Alternative BAM implementation

# 1 Survey Information

Table 1: Latinobarómetro "Bridge" Questions' Response Rates

| Stimulus | Respondents (N) | Response Rate (%) |
|---|---|---|
| José Mujica | 3557 | 17.61 |
| Juan Manuel Santos | 3558 | 17.61 |
| Sebastián Piñera | 4861 | 24.06 |
| Fernando Lugo | 4910 | 24.30 |
| Daniel Ortega | 5167 | 25.57 |
| Rafael Correa | 5887 | 29.14 |
| Felipe Calderón | 6198 | 30.68 |
| Alan García | 6396 | 31.66 |
| Cristina Kirchner | 8196 | 40.57 |
| Evo Morales | 10032 | 49.65 |
| "Lula" da Silva | 11138 | 55.13 |
| Barack Obama | 12762 | 63.17 |
| Fidel Castro | 12872 | 63.71 |
| Hugo Chávez | 13258 | 65.62 |

Table 2: PELA Survey "Bridge" Questions' Response Rates

| Stimulus | Respondents (N) | Response Rate (%) |
|---|---|---|
| Cristina Kirchner | 681 | 82.74 |
| Álvaro Uribe | 725 | 88.09 |
| "Lula" da Silva | 727 | 88.35 |
| Felipe Calderón | 798 | 96.96 |
| Hugo Chávez | 806 | 97.93 |
| Evo Morales | 808 | 98.17 |
| Barack Obama | 810 | 98.42 |

Table 3: Respondents, "Bridge" Questions, and Dates for PELA and CSES surveys

| Country | Respondents (N) | Bridge Questions | PELA Date | CSES Date |
|---|---|---|---|---|
| Brazil | 2129 | 9 | Apr.-Jun. 2010 | Nov. 2010 |
| Chile | 1286 | 7 | Jun.-Jul. 2010 | Dec. 2009-Jan. 2010 |
| Mexico | 2498 | 8 | Aug.-Dec. 2010 | Sept.-Dec. 2009 |
| Peru | 1663 | 10 | Oct.-Nov. 2011 | May 2011 |

Table 4: Merged CSES/PELA "Bridge" Questions and Response Rates

| Stimulus | Respondents (N) | Response Rate (%) |
|---|---|---|
| Brazil | | |
| Self | 1232 | 57.87 |
| Rousseff | 1314 | 61.72 |
| Serra | 1302 | 61.16 |
| PT | 1301 | 61.11 |
| PMDB | 1180 | 55.43 |
| DEM | 949 | 44.57 |
| PTB | 979 | 45.98 |
| PSDB | 1220 | 57.30 |
| PDT | 1009 | 47.39 |
| Chile | | |
| Self | 1004 | 78.07 |
| Pinera | 1197 | 93.08 |
| PDC | 1135 | 88.26 |
| PPD | 1116 | 86.78 |
| RN | 1141 | 88.72 |
| UDI | 1127 | 87.64 |
| PS | 1135 | 88.26 |
| Mexico | | |
| Self | 2153 | 86.19 |
| Calderon | 2068 | 82.79 |
| PRI | 2096 | 83.91 |
| PAN | 2080 | 83.27 |
| PRD | 2072 | 82.95 |
| Verde | 1926 | 77.10 |
| PT | 1784 | 71.42 |
| Panal | 1690 | 67.65 |
| Peru | | |
| Self | 1230 | 73.96 |
| Toledo | 1226 | 73.72 |
| Humala | 1260 | 75.77 |
| Fujimori | 1243 | 74.74 |
| Alianza Cambio | 1191 | 71.62 |
| APRA | 1213 | 72.94 |
| Solidaridad | 1210 | 72.76 |
| Fuerza 2011 | 1237 | 74.38 |
| Gana Peru | 1243 | 74.74 |
| Peru Posible | 1224 | 73.60 |

# 2 Selection-corrected estimates

To obtain selection-corrected estimates of the relationship between issues and ideology, I model the probability of being in the restricted sample and then use this information to examine the outcome of interest. In the first stage, the following probit regression is used:

$$P[y_i = 1|X_i] = \Phi(X_i\gamma)$$

where $y_i$ is: (1) a dummy variable, which takes a value of 1 for those who responded to the ideology question and 0 for those who did not respond (combining those who offered no response at all and those who stated they do not know); or (2) a dummy variable, which takes a value of 1 for those who rated 3 of more stimuli and 0 for those who failed to do it.

The vector $X_i$ of explanatory variables includes measures of the respondent's level of education, political interest, and political knowledge. I also control for the respondent's age, gender, and socioeconomic level (see below). Finally, $\gamma$ is a vector of parameters to be estimated; and $\Phi$ is the cumulative distribution function of the standard normal distribution. I include country fixed effects in all the models. To address heteroskedasticity, I employ Huber-White Robust standard errors (clustered at the country level). Table 5 presents the results (standard errors are in parentheses. $*$ indicates significance at a 10% level; $**$ indicates significance at a 5% level; $***$ indicates significance at a 1% level).

Table 5: Selection Equation – Correlates of Ideology

| | Self-Place Ideology | Estimated Ideology |
|---|---|---|
| Male | 0.202*** | 0.343*** |
| | (0.019) | (0.032) |
| Age | -0.005*** | -0.005*** |
| | (0.001) | (0.001) |
| Partisan Attachment | 0.584*** | 0.159*** |
| | (0.100) | (0.050) |
| Interest in Politics | 0.194*** | 0.111*** |
| | (0.026) | (0.021) |
| Understand Politics | 0.219*** | 0.229*** |
| | (0.044) | (0.024) |
| Attention to Politics | 0.042*** | 0.058*** |
| | (0.010) | (0.009) |
| Income Level | 0.150*** | 0.268*** |
| | (0.043) | (0.025) |
| College Education | 0.191*** | 0.646*** |
| | (0.061) | (0.053) |
| Constant | 0.619*** | -1.476*** |
| | (0.134) | (0.092) |
| Rho | 0.393 | -0.195** |
| | (0.557) | (0.076) |
| Sigma | -0.021 | -1.757*** |
| | (0.083) | (0.034) |
| Censored Obs. | 4,380 | 7,666 |
| Uncensored Obs. | 11,112 | 8,899 |
| Country FE | Yes | Yes |

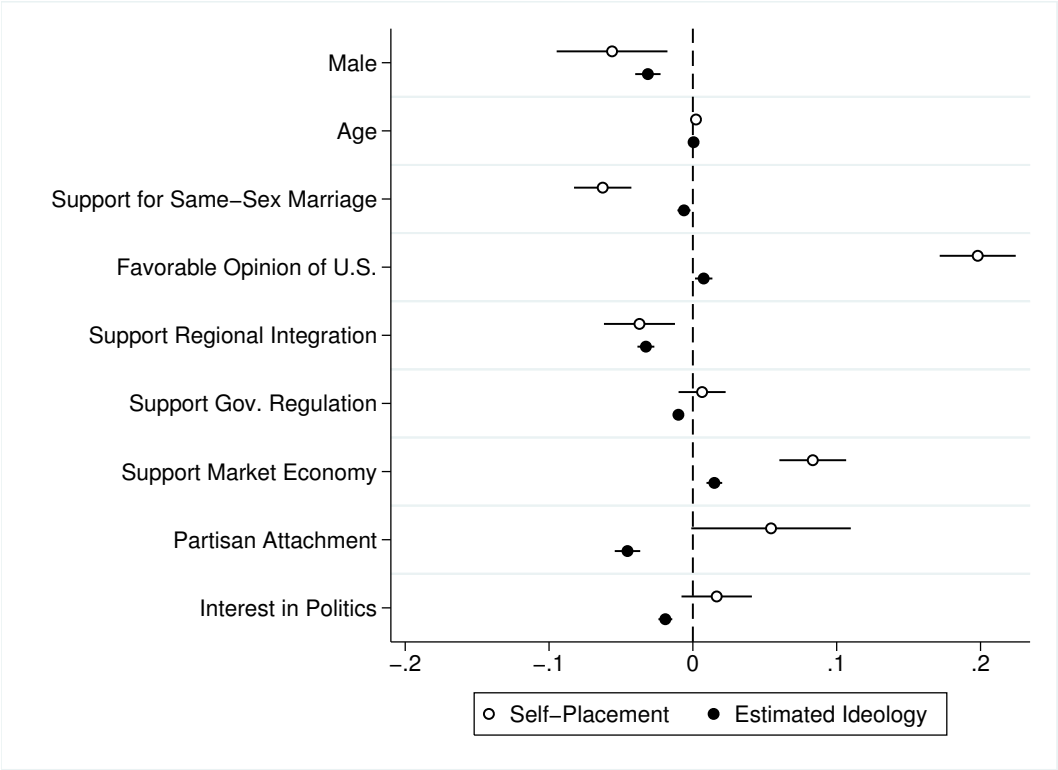
The second stage model takes the following form:

$$w_i^* = V_i'\delta + u_i$$

where $w_i^*$ denotes either: (1) the respondent's placement on the left-right scale, recoded here as z-scores; or (2) my recovered estimates of ideology. To assess the substantive determinants of the respondents' ideological positions ($V_i$), I also follow Zeichmeister and Corral (2013) and include measures capturing four issue dimensions that are relevant to Latin American politics: gay rights, relationship with the US, regional integration, and state-market relations. Lastly, $\delta$ is a vector of parameters to be estimated; and $u_i$ is an error term.

Note that $w_i^*$ is not observed if the respondent is not included in the restricted sample. The conditional expectation of $w_i$ is then given by $E[w_i|V_i, y_i = 1] = V_i\delta + E[u_i|V_i, y_i = 1]$. Under the assumption that the error terms are jointly normal, we obtain $E[w_i|V_i, y_i = 1] = V_i\delta + \rho\sigma_u\lambda(X_i\gamma)$, where $\rho$ is the correlation between the unobserved determinants of the propensity to be in the restricted sample ($\epsilon$) and the unobserved determinants of $w_i$; $\sigma_u$ is the standard deviation of $u$; and $\lambda$ is the inverse Mills ratio evaluated at $X_i\gamma$ (because the selection equation is estimated using a probit regression, $\sigma_\epsilon$ is assumed to be 1). The two sets of models can be jontly estimated using full maximum likelihood and including $\lambda$ as an additional explanatory variable in their second stage equations. Figure 1 presents a "ropeladder" plot of the selection-corrected estimates. Dots are point estimates and the spikes depict 95% confidence intervals.

Figure 1: Correlates of Ideology (selection-corrected estimates)



To obtain selection-corrected estimates of voters' ideological location (Figure 3 in the paper), I rely again on a Heckman selection model, and adopt the same specifictaions as the ones discussed above. In this case, to produce the selection variable, I estimate a probit model where the dependent variable takes a value of 1 for those who named a specific party as their vote choice and 0 for those who answered "Do not Know" or "Will not vote." Table 6 presents the results (standard errors are in parentheses. $*$ indicates significance at a 10% level; $**$ indicates significance at a 5% level; $***$ indicates significance at a 1% level).

Table 6: Selection Equation – Ideology of Representative Voter

| | States Vote Choice |
|---|---|
| Male | 0.223*** |
| | (0.023) |
| Age | -0.003** |
| | (0.001) |
| Partisan Attachment | 1.188*** |
| | (0.100) |
| Interest in Politics | 0.126*** |
| | (0.026) |
| Understand Politics | 0.169*** |
| | (0.030) |
| Attention to Politics | 0.042*** |
| | (0.008) |
| Income Level | 0.157*** |
| | (0.021) |
| College Education | 0.285*** |
| | (0.063) |
| Constant | -1.888*** |
| | (0.134) |
| Rho | -0.516** |
| | (0.239) |
| Sigma | -1.645*** |
| | (0.070) |
| Censored Obs. | 12,880 |
| Uncensored Obs. | 4,617 |
| Country FE | Yes |

Next, I use this information to correct the estimates of the correlation between issues and estimated ideology. Finally, I estimate a model using the respondents' ideology as the dependent variable and the selection instrument and country-specific constants as my independent variables:

$$IDEOLOGY_i = \sum_{1=1}^{I} \alpha_i COUNTRY_i + \beta\lambda_i + v_i,$$

where $I$ is an index identifying each country, the $\alpha$s are the country specific constants, $\lambda$ is the selection instrument, and $v_i$ is the error term. I partition the sample into two groups of respondents (based on their vote choice), and estimate a separate model for each group by

ordinary least squares regression. I thus generate two sets of selection-corrected parameters, one for repondents who support the government and the other for those in the opposition. The coefficients of the country-specific constants indicate the ideological location of each country's representative government/opposition voter.

**Details on variables used to estimate the selection equations**

1. *Partisan Attachment* is a dummy variable based on Q32N in the 2010 *Latinobarómetro* survey: "Is there any Political Party to which you feel closer to than the rest of the parties." Possible answers are: Yes (1), or No (0).

2. *Interest in Politics* is based on Q23STin the 2010 *Latinobarómetro* survey: "How interested are you in politics?". Possible answers are: Not at all interested (1); Few interested (2); Some interested (3); and Very interested (4).

3. *Understand Politics* is a dummy variable based on Q26STin the 2010 *Latinobarómetro* survey: "Some people say that politics is so complicated that people like us often cannot understand what is going on. Others are of the opinion that it is not so complicated and that they can understand what is going on. Which statement is closest to your way of thinking?" Possible answers are: Politics is not so complicated, (1) or Politics is so complicated (0).

4. *Attention to Politics* is a variable based on Q78ST/N in the 2010 *Latinobarómetro* survey: "How many days in the last week you look political news on TV?".

5. *Income Level* is a variable based on item S28 in the 2010 *Latinobarómetro* survey: "Assessment of the interviee's socio-economic level. Take as reference: quality of dwelling, quality of furniture, and the interviee's general appearance. Possible answers are: Very Bad (1); Bad (2); Average (3); Good (4); Very Good (5).

6. *College Education* is a dummy variable based on S14 in the 2010 *Latinobarómetro* survey: " What level of education do you have?" Coded as college education (1), less than college (0).

**Details on variables used to estimate the outcome equations**

1. *Support for Same-Sex Marriage* is based on Q14ST/N in the 2010 *Latinobarómetro* survey: "Views on same-sex marriage." Possible answers are: Strongly Disagree (1); Disagree (2); Agree (3); Strongly Agree (4).

2. *Favorable Opinion of United States* is based on Q39ST.A in the 2010 *Latinobarómetro* survey:"I would like to know your opinion about the United States." Possible answers are: Very Bad (1); Bad (2); Good (3); Very Good (4).

3. *Support Regional Integration* is based on Q43ST.A in the 2010 *Latinobarómetro* survey: "Generally speaking, are you very in favor, quite in favor, slightly against or very against of the economic integration of the countries of Latin America?" Possible answers are: Very Against (1); Slightly Against (2); Quite in Favor (3); Very Much in Favor (4).

4. *Support Government Regulation.* is based on Q63ST in the 2010 *Latinobarómetro* survey: " It is said that the state can solve the problems of our society because it has the means to do it. Would you say that the state can solve ?". Possible Answers are: The State can not solve any problem (0); Only a few problems (1); Enough problems (2); Most problems (3); All the problems (4).

5. *Support Market Economy* is based on Q75ST.A in the 2010 *Latinobarómetro* survey: "Private enterprise is indispensable to the development of the country" Possible Answers are: Strongly Disagree (1); Disagree (2); Agree (3); Strongly Agree (4).

6. *Partisan Attachment* is a dummy variable based on Q32N in the 2010 *Latinobarómetro* survey: "Is there any Political Party to which you feel closer to than the rest of the parties." Possible answers are: Yes (1), or No (0).

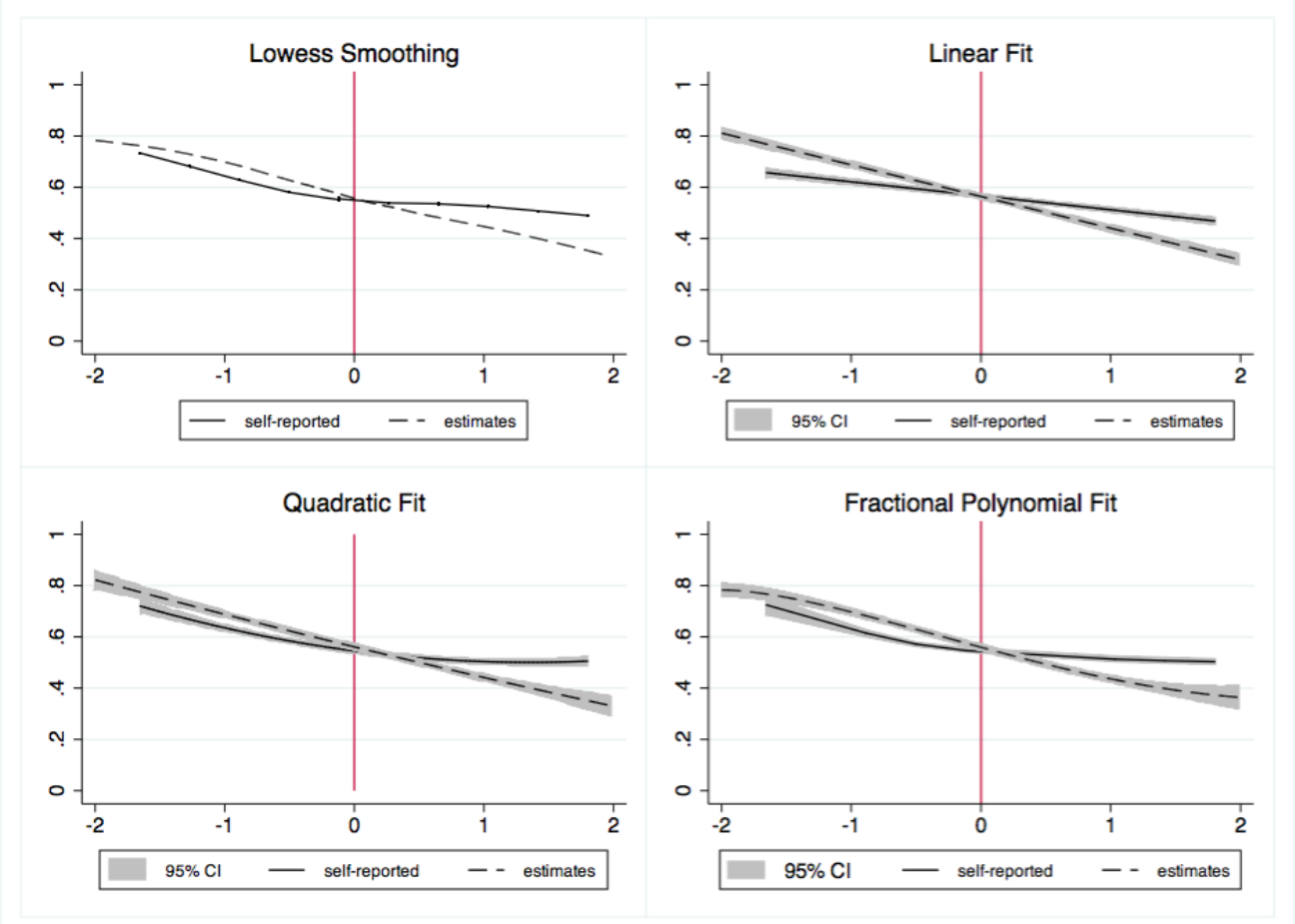
7. *Interest in Politics* is based on Q23STin the 2010 *Latinobarómetro* survey: "How interested are you in politics?". Possible answers are: Not at all interested (1); Few interested (2); Some interested (3); and Very interested (4).

# 3 Probability of voting for the government's party

The evidence presented in Figure 4 of the paper indicates that for the Latin American region as a whole, the nonparametric fit line for the recovered estimates of ideology is steeper than the one for the self-reported ideology. There are no confidence intervals associated with the fit lines. it is still possible, however, to establish if the difference in slope is statisticaly distinguishable from zero.

1 Figure 2 shows the relationship between a respondent's ideology and the probability of voting for the government's party using locally weighted, linear, quadratic and polynomial regression for the 18 Latin American countries included in the analysis.

Figure 2: Expected probability of voting for government




The results suggest that the difference in the fit lines' slopes is statisticaly significant.

2 An alternative approach is to compare the quantiles of the distributions of the two non-parametric estimates using a Harrell-Davis estimator in conjunction with a percentile bootstrap. The WRS package on CRAN contains a function to compare user-defined quantiles for two samples using a HarrellDavis estimator in conjunction with a percentile bootstrap. Exact tests are discussed in Li, Tiwari and Wells (1996) and Kosorok (1999); and the computational aspects are described in Wilcox et. al (2014). Table 7 shows how the two distributions differ in the requested quantiles, as well as the confidence intervals for each quantile, when the data for the 18 Latin American countries included in the analysis are considered.

Table 7: Differences between the two distributions

| quantile | difference | c.i.(low) | c.i.(up) | crtitical p-value | p-value |
|---|---|---|---|---|---|
| 0.05 | 0.095 | 0.091 | 0.099 | 0.025 | 0.000 |
| 0.25 | 0.047 | 00.044 | 0.051 | 0.016 | 0.000 |
| 0.5 | -0.001 | -0.006 | 0.004 | 0.05 | 0.701 |
| 0.75 | -0.0781 | -0.083 | -0.072 | 0.012 | 0.000 |
| 0.95 | -0.056 | -0.077 | -0.027 | 0.01 | 0.000 |

The column labeled "p-value" shows the p-value for a single quantile bootstrapping test. Because multiple tests are performed (one for each quantile), the overall Type 1 error (defaulting to .05) is controlled by the Hochberg method. Therefore, for each p-value, a critical p-value that must be undercut, is calculated. The results indicate that there are significant differences between the two distributions at the .05, .25, .75 and .95 quantiles. Figure 3 provides a graphical representation of the results.

Figure 3: Differences between the distributions of the two non-parametric estimates

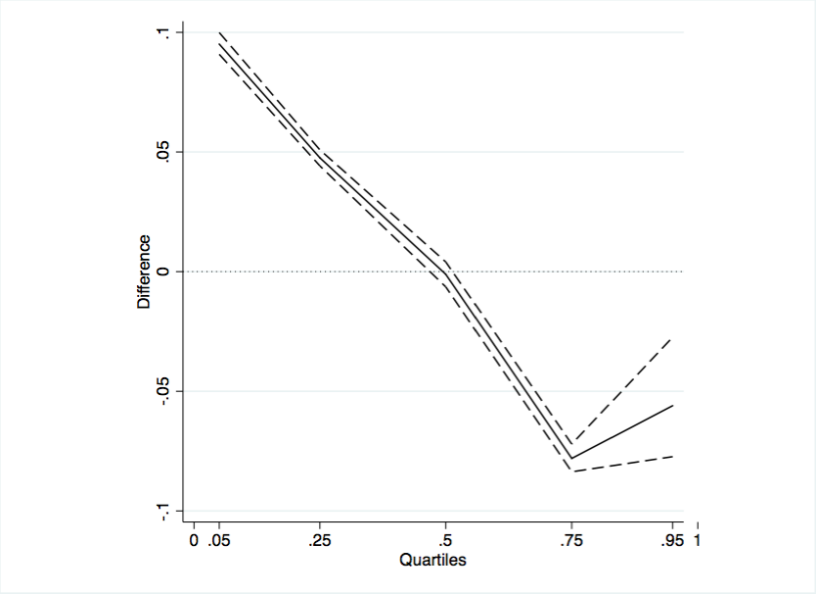
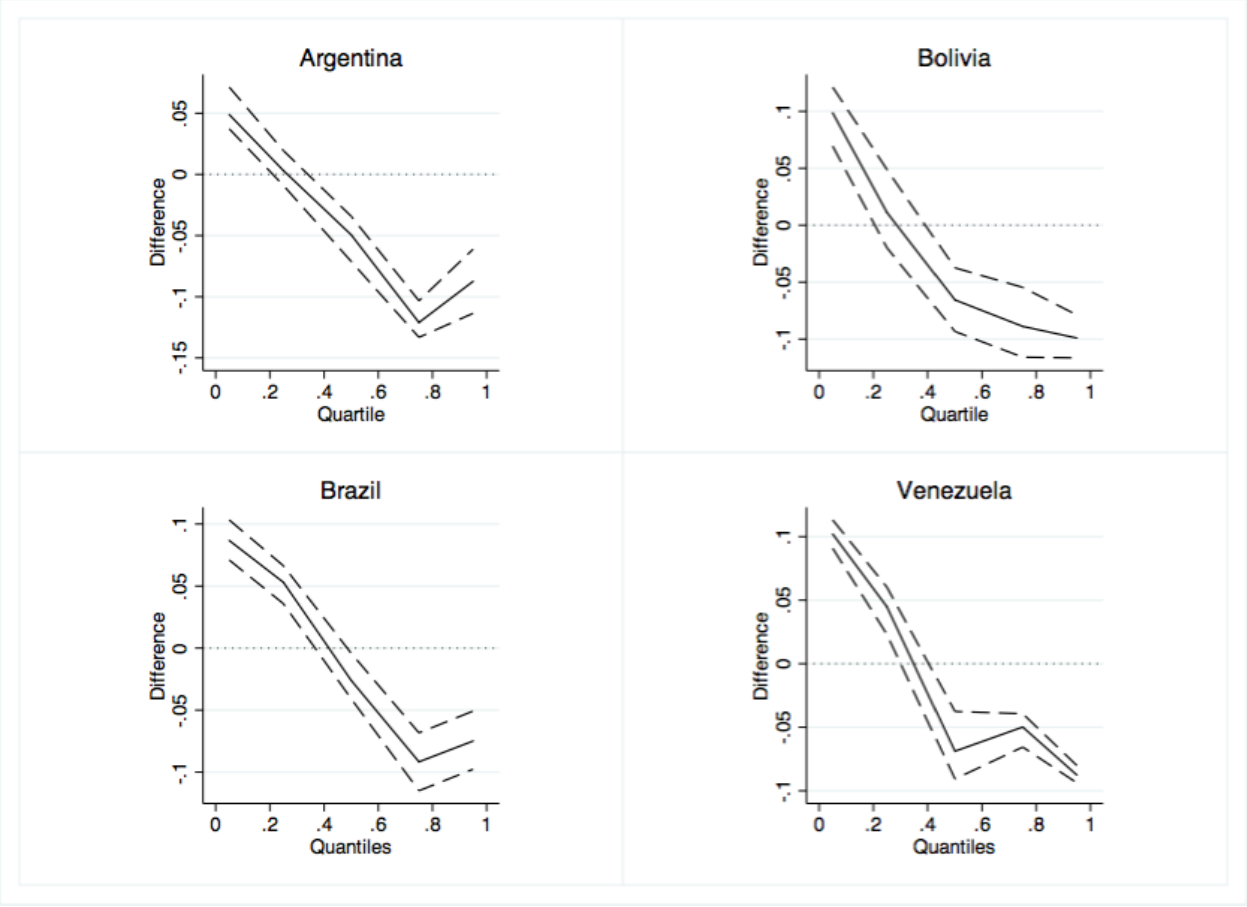


Figure 4 shows a graphical representation of the results for the cases of Argentina, Bolivia, Brazil, and Venezuela.

Figure 4: Differences between the distributions of the two non-parametric estimates



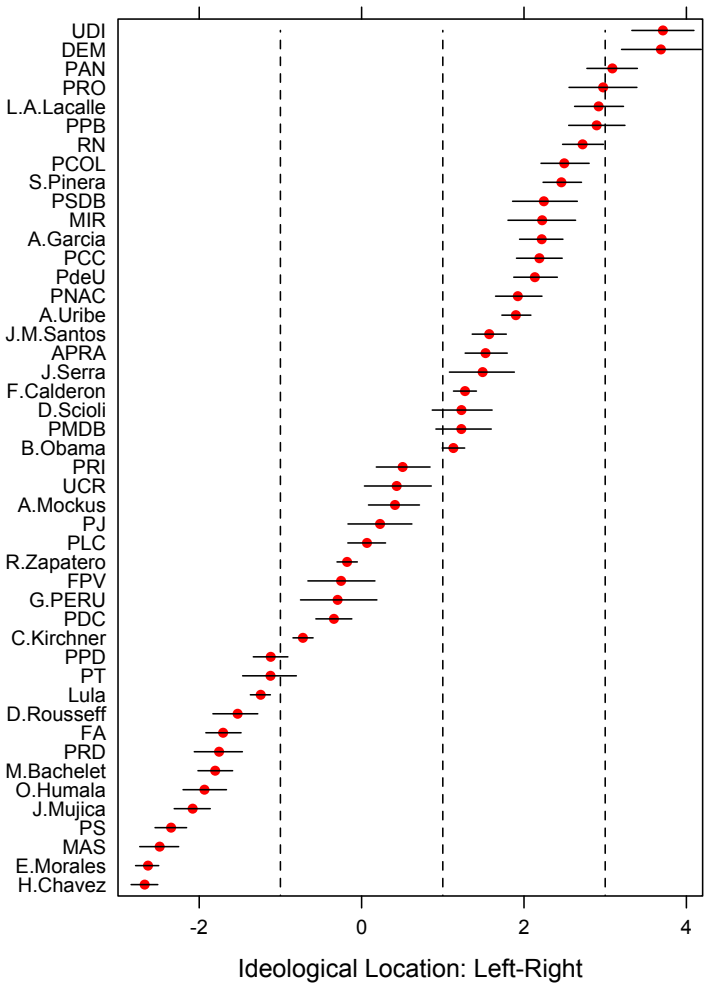
# 4 Alternative BAM implementation

In the alterantive BAM implementation, I assume that the perceived location of stimulus $j$ by individual $i$ follows a distribution:

$$z_{ij} \sim N(\mu_{ij}, \tau_j)$$
$$\mu_{ij} = \alpha_i + \beta_i Z_j$$

I employ non-informative conjugate priors $\alpha_i \sim N(0,1)$, $\beta_i \sim N(0,1)$, $\tau_j \sim Gamma(1, 0.1)$, and $Z_j \sim N(0,1)$. Identification is obtained by constraining Hugo Chávez to be negative. Figure 5 presents the point estimates and 95% credible intervals for the stimuli (arranged ideologically from left to right).

Figure 5: Stimuli Location: point estimates and 95% credible intervals



# References

Kosorok Michael R., and Chin-Yu Lin. 1999. "The Versatility of Function-Indexed Weighted Log-Rank Statistics," *Journal of the American Statistical Association*, Vol. 94: 320-332.

Li, Gang, Ram C. Tiwari and Martin T. Wells. 1996. "Quantile Comparison Functions in Two-Sample Problems, With Application to Comparisons of Diagnostic Markers," *Journal of the American Statistical Association*, Vol. 91: 689-698.

Wilcox, Rand R., David M. Erceg-Hurn , Florence Clark , and Michael Carlson. 2014. "Comparing two independent groups via the lower and upper quantiles," *Journal of Statistical Computation and Simulation*, Vol. 84: 1543-1551.