# Online Appendix: Supporting Information for "Non-separable Preferences in the Statistical Analysis of Roll Call Votes"


## Garret Binding[1] and Lukas F. Stoetzer[2]

[1] *University of Zurich. Email: binding@ipz.uzh.ch*
[2] *Humboldt University of Berlin. Email: lukas.stoetzer@hu-berlin.de*


## 1 Stan Code for Non-separable IRT Model

### 1.1 Monte Carlo

```
functions {
   real pow_sum (vector x, real z){
    real y = 0;

    for (i in 1:num_elements(x)){
      y = y + x[i]^z;
    }
    return y;
   }

  real sqsum (vector x){
    real y = 0;

    for (i in 1:num_elements(x)){
      y = y + x[i]*x[i];
    }
    return y;
   }

   real calc_cor (vector x, vector y){
     real r;
     r = sum((x - mean(x)).*(y - mean(y))) / sqrt(sqsum(x - mean(x))*sqsum(y - mean(y)));
     return r;
   }
}
data {
  int<lower=1> I;               // proposals
  int<lower=1> J;               // legislators
  int<lower=1> D;               // number of latent dimensions
  int<lower=1> K;               // number of exp. variables
  int<lower=0,upper=1> Y[J,I];  // binary y
  real theta_in[J,K,D];         // explanatory variables for ideal points
  matrix[J,D] theta_true;
  int<lower=0,upper=1> incl_nonsep;  // indicator to use separable or nonseparable model
  int<lower=0,upper=1> incl_logl;  // indicator to calculate log likelihood
}
parameters {
  ordered[D] sal_free; // salience of dimensions - free
  real<lower=0> tau; // scale of weights matrix overall
  real<lower=0> tau_int; // scale of weights matrix internally
  matrix[J,D] theta_free; // free ideal points
  matrix[I,D] bk_free; // free discrimination
  vector[I] ak; // difficulty
  cholesky_factor_corr[D] L; // correlation component of weight matrix
  row_vector<lower=0>[K] gamma[D]; // coefficients for linear regression on ideal points, lower bound to prevent switching
}
transformed parameters {
  corr_matrix[D] DL; // correlation component of weight matrix
  cov_matrix[D] weights; // final weight matrix
  matrix[J,D] theta_t;
  matrix[I,D] bk; // standardized discrimination (var 1)
  matrix[J,D] theta; // standardized ideal points (mean 0, var 1)
  simplex[D] sal = softmax(sort_desc(sal_free)); // salience of dimensions - rescaled to simplex
  DL = L * L';

  {
    vector[D] sd_theta;
    vector[D] sd_bk;
    vector[D] mean_theta;

    // needs this constraint of sum(abs(...)) == 1 here so that weights is identified internally.
    // otherwise, same variance can be gotten by varying either var or cov within the final matrix
    if (incl_nonsep == 1){
      weights = quad_form_diag(DL, sqrt(sal * tau_int)) / sum(fabs(to_vector(quad_form_diag(DL, sqrt(sal * tau_int))))) * tau^2;
    } else {
      weights = quad_form_diag(diag_matrix(rep_vector(1.00, D)), sqrt(sal * tau));
    }
    for (d in 1:D){
      theta_t[,d] = to_matrix(theta_in[,,d]) * to_vector(gamma[d]) + theta_free[,d];
    }

    for (d in 1:D){
```


***Political Analysis (2022)***

**DOI:** 10.1017/pan.xxxx.xx



**Corresponding author**
Garret Binding



**Edited by**
John Doe

```
      sd_theta[d] = sd(theta_t[,d]);
      sd_bk[d] = sd(bk_free[,d]);
      mean_theta[d] = mean(theta_t[,d]);
    }

    for (d in 1:D){
      theta[,d] = (theta_t[,d] - mean_theta[d]) / sd_theta[d];
      bk[,d] = (bk_free[,d]) / sd_bk[d];
    }
  }
}
model {
  to_vector(bk_free) ~ std_normal();
  ak ~ std_normal();
  sal_free ~ std_normal();
  tau ~ std_normal();
  tau_int ~ normal(D*1.00, 1); // in expectation, the 'internal' weights part has a variance of 2 along the diagonals, before being rescaled by tau
  to_vector(theta_free) ~ std_normal();
  for (d in 1:D){
    gamma[d] ~ normal(1, 1);
  }
  L ~ lkj_corr_cholesky(1); // uniform dist

  {
    matrix[J,I] ystar;
    ystar = (theta * weights * bk');
    for (i in 1:I){
      Y[,i] ~ bernoulli(Phi_approx(ystar[,i] + ak[i]));
    }
  }

}
generated quantities{
  vector[D] r_sq; // share of var in final theta explained by lpm
  vector[D] r_true; // correlation of final theta with true theta
  real r_dim; // correlation of final theta with true theta
  real m_prop_compl; // mean proposal complexity
  real sd_prop_compl; // sd proposal complexity
  real p_share; // share of correct votes
  matrix[J,D] theta_pred; // predicted values based on lpm
  matrix[J,D] theta_resid; // residual values
  row_vector[K] gamma_post[D]; // coef from lpm on rescaled theta
  vector[I*J*incl_logl] log_lik;

  r_dim = calc_cor(theta[,1], theta[,2]);

  for (d in 1:D){
    theta_pred[,d] = to_matrix(theta_in[,,d]) * to_vector(gamma[d]);
    theta_resid[,d] = (theta_pred[,d] - mean(theta_pred[,d])) - (theta_t[,d] - mean(theta_t[,d]));
    r_sq[d] = sd(theta_pred[,d]) * sd(theta_pred[,d]) / (sd(theta_pred[,d]) * sd(theta_pred[,d]) + sd(theta_resid[,d]) * sd(theta_resid[,d]));
    r_true[d] = calc_cor(theta_true[,d], theta[,d]);
    gamma_post[d] = to_row_vector((to_matrix(theta_in[,,d])' * to_matrix(theta_in[,,d])) \ to_matrix(theta_in[,,d])' * theta[,d]);
  }
  {
    vector[I] prop_compl;
    for (i in 1:I) prop_compl[i] = (pow_sum(bk[i,]', 2.00))^2 / pow_sum(bk[i,]', 4.00); // Hofmann, R. J. (1977). Indices descriptive of factor complexity. Journal of General
    m_prop_compl = mean(prop_compl);
    sd_prop_compl = sd(prop_compl);
  }
  if (incl_logl == 1){
    {
      matrix[J,I] ystar;
      real pred = 0.0;
      for (i in 1:I){
        for (j in 1:J){
         ystar[j,i] = bernoulli_lpmf(Y[j,i] | Phi_approx((theta[j,] * weights * bk[i,]') + ak[i]));
         if (Y[j,i] == bernoulli_rng(Phi_approx((theta[j,] * weights * bk[i,]') + ak[i]))){
           pred += 1.00;
         }
        }
      }
      log_lik = to_vector(ystar);
      p_share = pred / (I*J*1.00);
    }
  }
}
```

## 1.2 Applications

```
functions {
    real sqsum (vector x){
    real y = 0;

    for (i in 1:num_elements(x)){
      y = y + x[i]*x[i];
    }
    return y;
   }

  real pow_sum (vector x, real z){
    real y = 0;

    for (i in 1:num_elements(x)){
      y = y + x[i]^z;
    }
```

```
    return y;
   }

   real calc_cor (vector x, vector y){
     real r;
     r = sum((x - mean(x)).*(y - mean(y))) / sqrt(sqsum(x - mean(x))*sqsum(y - mean(y)));
     return r;
   }
}
data {
  int<lower=1> I;               // proposals
  int<lower=1> J;               // legislators
  int<lower=1> N;               // nr of observed rc
  int<lower=1> J_old;           // legislators
  int<lower=1> D;               // number of latent dimensions
  int<lower=1> K;               // number of exp. variables
  int<lower=0,upper=1> T;       // include time component
  int<lower=-1,upper=1> Y[J,I]; // binary y
  int<lower=1,upper=I> ii[N];
  int<lower=1,upper=J> jj[N];
  real theta_in[J,K,D];         // explanatory variables for ideal points
  int<lower=0,upper=1> incl_nonsep;  // indicator to use separable or nonseparable model
  int<lower=1,upper=J> theta_old_ind[J_old];
  real theta_old[J_old,D];
}
parameters {
  simplex[D] sal; // salience of dimensions
  real<lower=0> tau; // scale of weights matrix overall
  real<lower=0> tau_int; // scale of salience internally in nonseparable setting
  matrix[J,D] theta_free; // free ideal points
  matrix[I,D] bk_free; // free discrimination
  vector[I] ak; // difficulty
  cholesky_factor_corr[D] L; // correlation component of weight matrix
  vector<lower=0>[D] gamma; // coefficients for linear regression on ideal points, lower bound to prevent switching
}
transformed parameters {
  corr_matrix[D] DL; // correlation component of weight matrix
  cov_matrix[D] weights; // final weight matrix
  matrix[J,D] theta_t;
  matrix[I,D] bk; // standardized discrimination (mean 0, var 1)
  matrix[J,D] theta; // standardized ideal points (var 1)

  DL = L * L';

  {
    vector[D] sd_theta;
    vector[D] sd_bk;
    vector[D] mean_theta;

    // it needs this constraint of sum(abs(...)) == 1 here so that weights is identified internally.
    // otherwise, same variance can be gotten by varying either var or cov within the final matrix
    if (incl_nonsep == 1){
      weights = quad_form_diag(DL, sqrt(sal * tau_int)) / sum(fabs(to_vector(quad_form_diag(DL, sqrt(sal * tau_int))))*tau^2;
    } else {
      weights = quad_form_diag(diag_matrix(rep_vector(1, D)), sqrt(sal * tau));
    }

    for (d in 1:D){
      theta_t[,d] = to_vector(theta_in[,1,d]) * (gamma[d]) + to_vector(theta_in[,2,d]) + theta_free[,d];
    }

    for (d in 1:D){
      sd_theta[d] = sd(theta_t[,d]);
      sd_bk[d] = sd(bk_free[,d]);
      mean_theta[d] = mean(theta_t[,d]);
    }

    for (d in 1:D){
      theta[,d] = (theta_t[,d] - mean_theta[d]) / sd_theta[d];
      bk[,d] = bk_free[,d] / sd_bk[d];
    }
  }
}
model {
  to_vector(bk_free) ~ std_normal();
  ak ~ std_normal();
  sal ~ dirichlet(rep_vector(1.00, D));
  tau ~ std_normal();
  tau_int ~ normal(D*1.00, 1); // in expectation, the 'internal' weights part has a variance of D along the diagonals, before being rescaled by tau
  to_vector(theta_free) ~ std_normal();
  for (d in 1:D){
    gamma[d] ~ normal(1, 1);
  }
  L ~ lkj_corr_cholesky(1); // uniform dist
  for (n in 1:N){
    Y[jj[n],ii[n]] ~ bernoulli(Phi_approx((theta[jj[n],] * weights * bk[ii[n],]') + ak[ii[n]]));
  }
}
generated quantities{
  vector[D] r_sq; // share of var in final theta explained by lpm
  vector[D] r_old; // correlation of final theta with theta at t-1
  real r_dim; // correlation of final theta across dims
  real m_prop_compl; // mean proposal complexity
  real sd_prop_compl; // sd proposal complexity
  real pred = 0.0;
  real p_share; // share of correct votes
  matrix[J,D] theta_pred;
```

```
  matrix[J,D] theta_resid;
  row_vector[K+T] gamma_post[D];
  vector[N] log_lik;

  r_dim = calc_cor(theta[,1], theta[,2]);

  for (d in 1:D){
    theta_pred[,d] = to_vector(theta_in[,1,d])*(gamma[d]) + to_vector(theta_in[,2,d]);
    theta_resid[,d] = (theta_pred[,d] - mean(theta_pred[,d])) - (theta_t[,d] - mean(theta_t[,d]));
    r_sq[d] = sd(theta_pred[,d])*sd(theta_pred[,d]) / (sd(theta_pred[,d])*sd(theta_pred[,d]) + sd(theta_resid[,d])*sd(theta_resid[,d]));

    if (T == 0){
      r_old[d] = 0.0;
    } else{
      r_old[d] = calc_cor(to_vector(theta_old[,d]), theta[theta_old_ind,d]);
    }
    if (T == 0){
      gamma_post[d] = to_row_vector((append_col(rep_vector(1,J),to_matrix(to_vector(theta_in[,1,d])))'*append_col(rep_vector(1,J),to_matrix(to_vector(theta_in[,1,d])))) \ app
    } else {
      gamma_post[d] = to_row_vector((append_col(rep_vector(1,J),to_matrix(theta_in[,,d]))'*append_col(rep_vector(1,J),to_matrix(theta_in[,,d]))) \ append_col(rep_vector(1,J),
    }

  }

  {
    vector[I] prop_compl;
    for (i in 1:I) prop_compl[i] = (pow_sum(bk[i,]', 2.00))^2 / pow_sum(bk[i,]', 4.00); // Hofmann, R. J. (1977). Indices descriptive of factor complexity. Journal of General
    m_prop_compl = mean(prop_compl);
    sd_prop_compl = sd(prop_compl);
  }

  for (n in 1:N){
    log_lik[n] = bernoulli_lpmf(Y[jj[n],ii[n]] | Phi_approx((theta[jj[n],] * weights * bk[ii[n],]') + ak[ii[n]]));
    if (Y[jj[n],ii[n]] == bernoulli_rng(Phi_approx((theta[jj[n],] * weights * bk[ii[n],]') + ak[ii[n]]))){
          pred += 1.00;
        }
  }
  p_share = pred / (I*J*1.00);
}
```

## 2 Monte Carlo Simulations

We describe the data generating process of the Monte Carlo simulations as well as additional results in this section.

### 2.1 Data Generating Process

The ideal points of 100 artificial legislators in two dimensions should (i) be a function of a systematic component and (ii) be allowed to correlate across dimensions. To achieve this, we first sample values at random from a standard multivariate normal distribution in two dimensions where the correlation between dimensions is set to values between -.8 and .8 (in steps of .1) at random. Then, these continuous values are transformed into group membership indicator variables by assigning values below the 1/3 percentile of each dimension-wise distribution a value of -1, those above the 2/3 percentile a 1, and those in between a 0. This corresponds to a scenario where we can (roughly) describe the substance of conflict along a dimension and classify political actors accordingly. The resulting vector of length 100 is $z_d$ for each dimension $d$ and has a variance of 2/3 in expectation.

Then, ideal points are sampled from a normal distribution $\boldsymbol{\theta_{.d}} \sim \mathcal{N}(\gamma_d z_d, \sqrt{1/3})$, where $\gamma_d = 1$. This means that roughly 2/3 of the variation in the resulting ideal points is due to the systematic component $z_d$ and roughly 1/3 of the variation is random. Finally, the ideal points are rescaled to a mean of 0 and a standard deviation of 1 on each dimension. This approach generates ideal points that are (i) a function of the systematic component $\hat{\theta}_{.d} = \gamma_d z_d$ and (ii) are correlated across dimensions at random (with a range of correlation ranging between approximately -.6 and .6).

The difficulty parameters $\boldsymbol{\alpha}$, the discrimination parameters $\boldsymbol{\beta}$, and an error term $\epsilon$ are drawn from a standard normal distribution, $\mathcal{N}(0, 1)$. Similar to the ideal points $\boldsymbol{\theta}$, the $\boldsymbol{\beta}$ parameters are rescaled to unit variance on each dimension. The difficulty parameters $\boldsymbol{\alpha}$ are multiplied by .2, so that the range of mean approval is between .15 and .85 (i.e., we exclude lopsided roll-call votes). The weight matrices $\mathbf{A}$ are decomposed into a correlation matrix and scaling parameters: $\mathbf{A} = \tau \boldsymbol{\Omega} \tau$ (Barnard, McCulloch, and Meng 2000). $\boldsymbol{\Omega}$ is a $2 \times 2$ correlation matrix in which the off-diagonal values are sampled from values between -.8 and .8 (in steps of .1) at random. The value of

$\tau_2$, the relative salience of the secondary dimension, is sampled at random from {.15, .25, .35, .45} (with equal probabilities) and $\tau_1 = 1 - \tau_2$. $\mathbf{A}$ is then rescaled so that the sum of the squared values of $\mathbf{A}$ equals 2, implying that roughly 2/3 of the variation on the underlying latent scale $y^*_{ij}$ across legislators $i$ and proposals $j$ is accounted for by the component $\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\theta}$ (as the variance of $\epsilon$ equals 1). Finally, Yea- and Nay-votes for roll calls $j$ and legislators $i$ are simulated from a binomial distribution with

$$P(y_{ij} = 1) = \Phi(y^*_{ij}) = \Phi(\boldsymbol{\beta}'_j\mathbf{A}\boldsymbol{\theta}_i + \alpha_j + \epsilon_{ij}).$$

## 2.2 Additional Results

We focus on the estimated weight matrices $\mathbf{A}$ in Figures 1 and 2, estimated ideal points $\boldsymbol{\theta}$ in Figures 3, 4, 5, and 6, and results from model comparison in Figures 7, 8, and 9. We consider results (i) across all scenarios, (ii) across different levels of salience of the primary dimension, and (iii) across different levels of correlation between the original ideal points (we distinguish the direction and magnitude of correlation by quintiles).

Figure 1 and Figure 2 focus on elements of the estimated weight matrices $\mathbf{A}$. Figure 1 mirrors the findings in the main text: the separable specification overestimates (underestimates) the relative salience of the primary (secondary) dimension at increasing absolute levels of non-separability. This pattern is more pronounced when the salience of both dimensions is more similar, while it is similar across different levels of correlation across dimensions of ideal points. Figure 2 presents the deviations in estimated non-separability values from the non-separable specification. Across all conditions considered, the extent of non-separability is slightly underestimated at higher values of non-separability. In relative terms, this underestimation is small compared to the range of possible values between -.8 and .8.

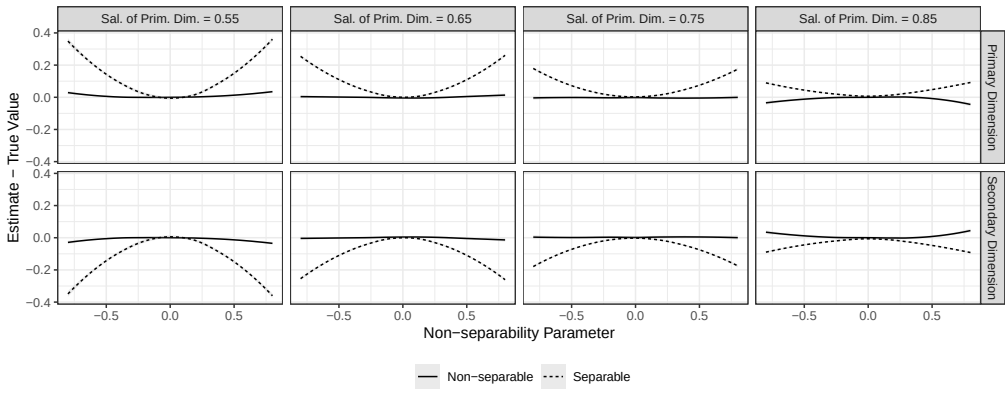
Figures 3, 4, 5 and 6 focus on the estimated ideal points $\boldsymbol{\theta}$. The correlation between estimated and true ideal points is shown in Figure 3, the $R^2$ statistic comparing predicted ideal points $\boldsymbol{\theta}^*$ to observed ideal points $\boldsymbol{\theta}$ is presented in Figure 4, and estimated $\gamma$ parameters are visualized in Figure 5. The deviations of the separable specification discussed in the main text are replicated at all levels of dimensional salience, but are most pronounced when both dimensions are similarly salient and / or when ideal points are not strongly correlated across dimensions. Finally, Figure 6 reproduces the findings in the main text: the separable specification accommodates non-separability in the decision making process of legislators by estimating correlated ideal points.

Figures 7, 8, and 9 focus on different approaches of model comparison. In Figure 7, we compare the performance of both model specifications via the LOO information criteria (Vehtari, Gelman, and Gabry 2017). Roughly, this pairwise comparison tells us whether the added complexity of allowing non-separability in the specification (and estimating an additional parameter) is worth it when considering the resulting log-likelihood of the estimated model compared to the separable specification. The resulting comparisons between the two specifications are shown in Figure 7. In Figure 7a (left panel), we show the share of pairwise comparisons in which either the separable or the non-separable specification performs better than the other one. When non-separability is low, the outcome of this comparison can go either ways, but as non-separability increases in absolute terms, the non-separable specification outperforms the separable specification consistently.

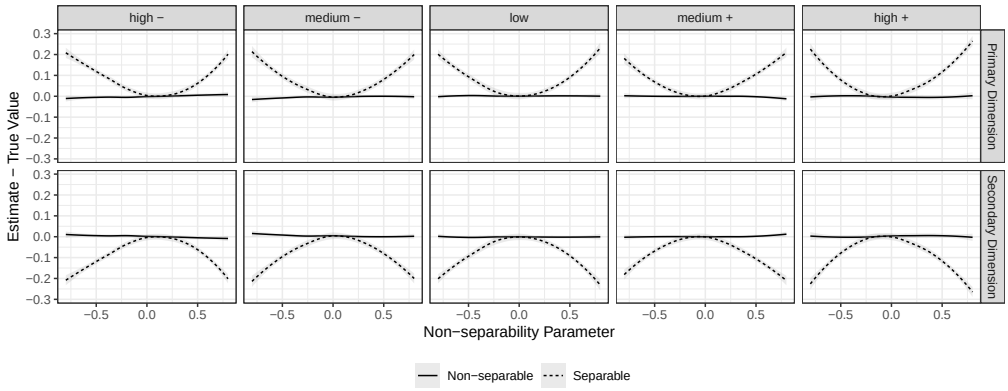
In Figure 7b (right panel), we present the share of comparisons for which this difference is significantly different from 0 by taking the standard error of the difference in means between the two into account. When non-separability is low, the difference between the two specifications is significant in around 30% of all comparisons (the sum of the two lines). As non-separability increases and the separable model performs consistently worse than the non-separable specification, this share decreases to around 20%. This implies that even if a non-separable specification generally performs better at high values of non-separability, the difference between the two need not necessarily be substantially significant when compared in this manner.

In Figure 8, we show the predictive accuracy of either model specification across all simulations (i.e., the share of all roll call votes predicted correctly). Both specifications perform equally well when compared by this metric. Because of the latent state of estimated parameters, different configurations of quantities can result in equal predictions–albeit at the cost of a distorted representation of political conflict.

In Figure 9, we consider the mean complexity of proposals' discrimination parameters across dimensions and simulations (Hofmann 1977). Lower values towards 1 indicate that proposals uniquely discriminate on one dimension, while higher values towards $D = 2$ indicate that proposals discriminate equally on both dimensions. Lower values would therefore imply that the discrimination structure of proposals across dimensions is simpler, with proposals discriminating on either dimension, but not on both. In our simulations, there is little difference between the two specifications concerning this quantity.

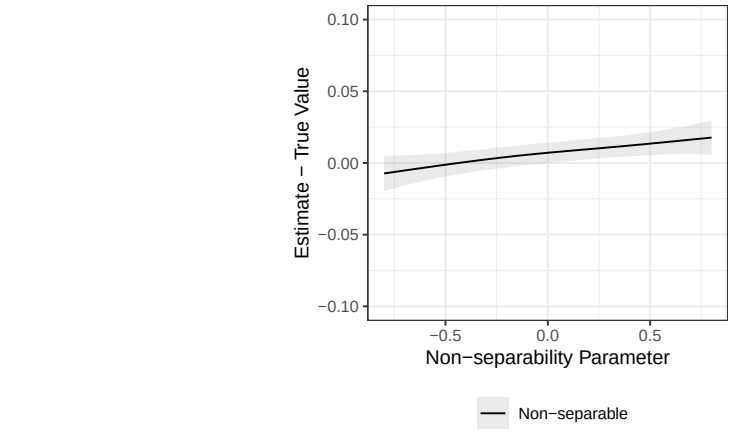



(a) Across salience of primary dimension.

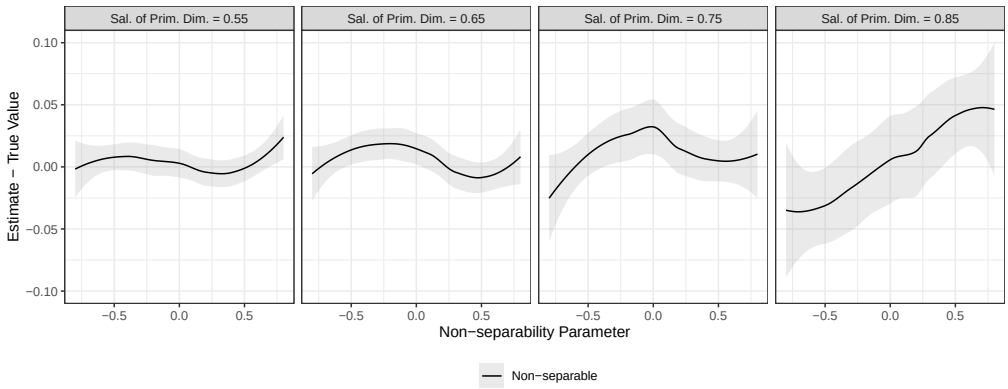



(b) Across correlation of true ideal points.

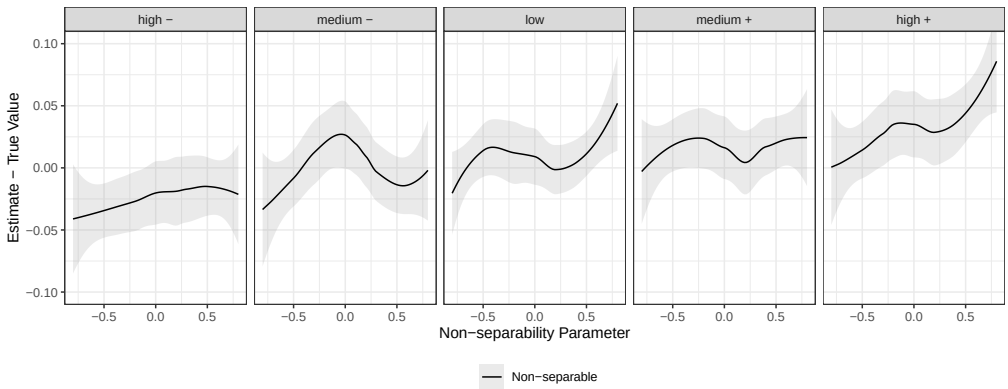
**Figure 1.** Deviation in estimated dimensional salience in weight matrices **A**.




(a) Across all scenarios.

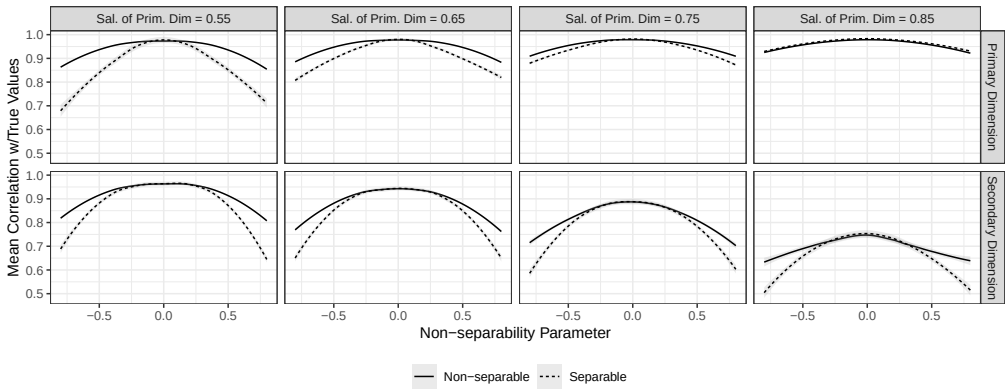



(b) Across salience of primary dimension.

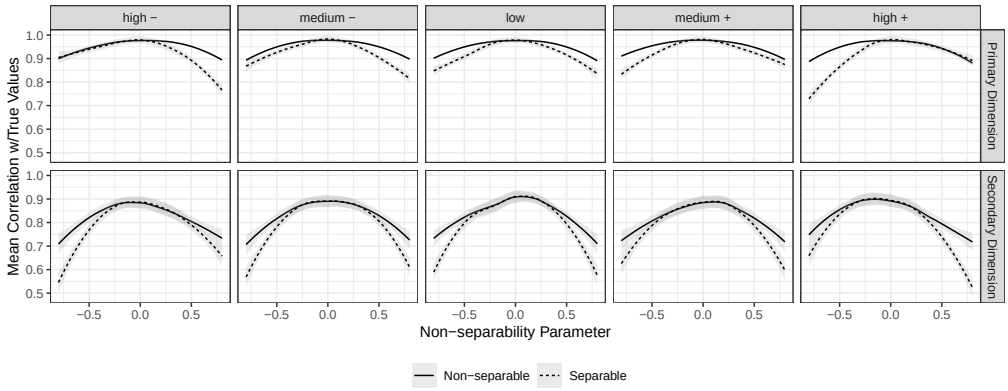



(c) Across correlation of true ideal points.

**Figure 2.** Deviation in estimated non-separability parameter in weight matrices **A**.

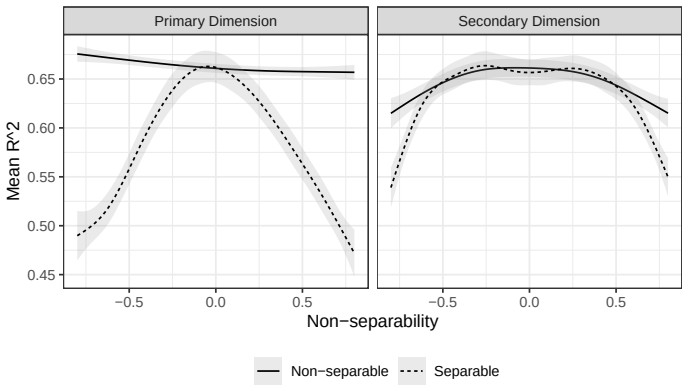



(a) Across salience of primary dimension.

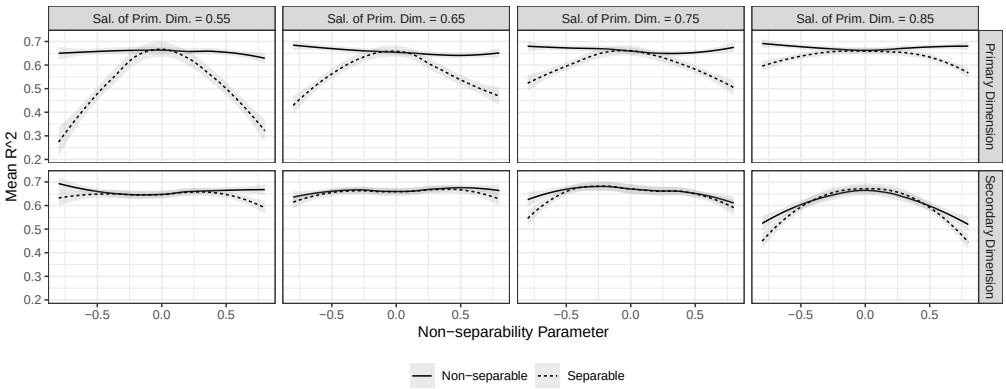



(b) Across correlation of true ideal points.

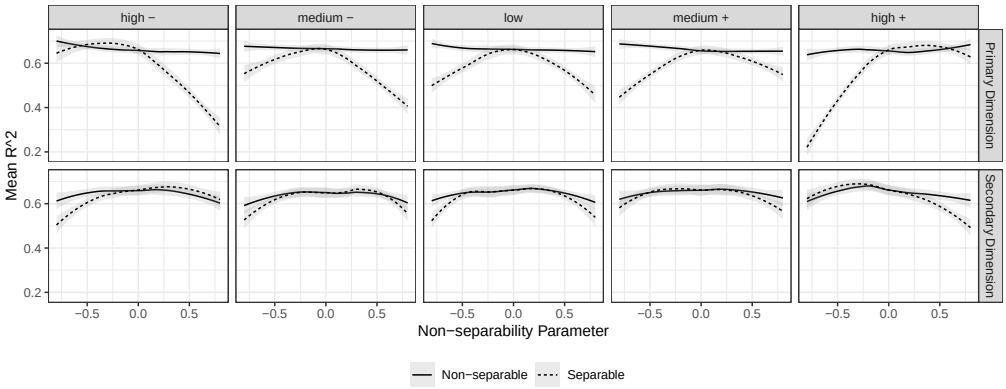
**Figure 3.** Correlation of estimated ideal points with true ideal points.




(a) Across all scenarios.

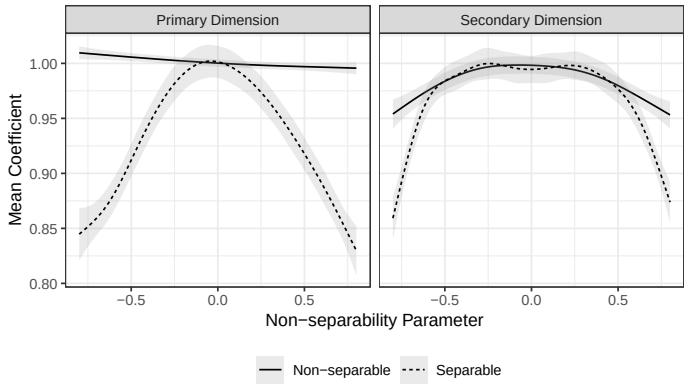



(b) Across salience of primary dimension.

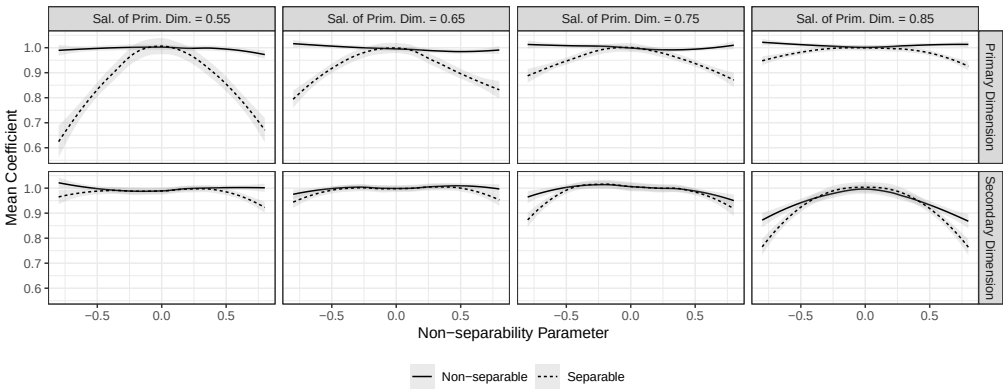



(c) Across correlation of true ideal points.

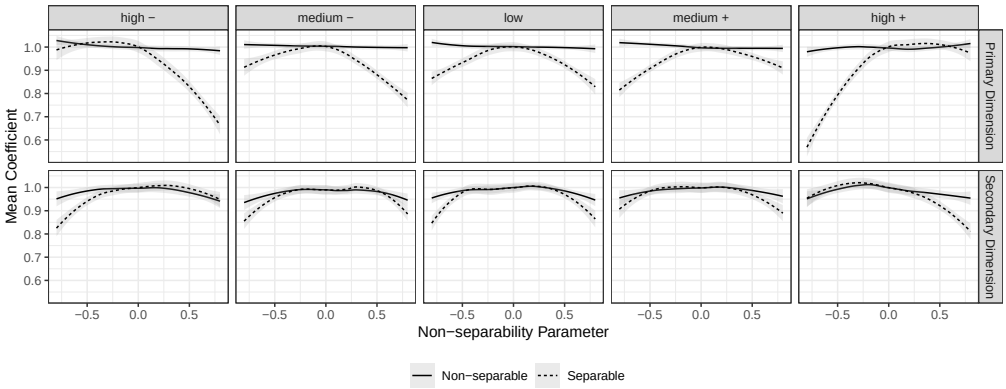
**Figure 4.** Estimated $R^2$.




(a) Across all scenarios.

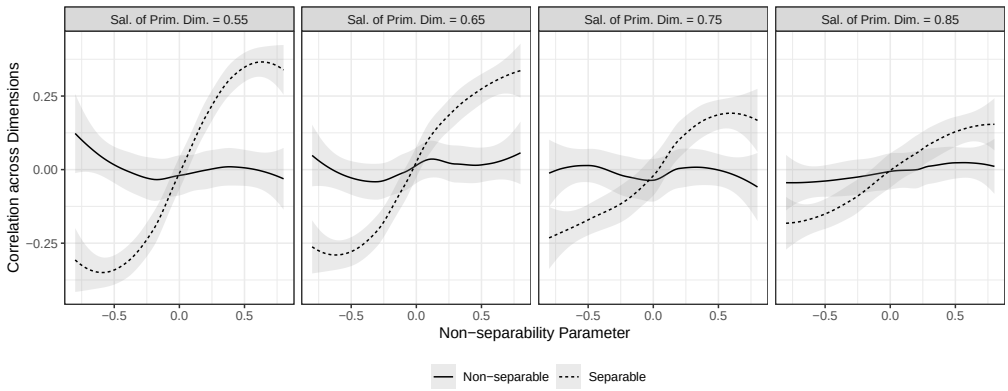



(b) Across salience of primary dimension.

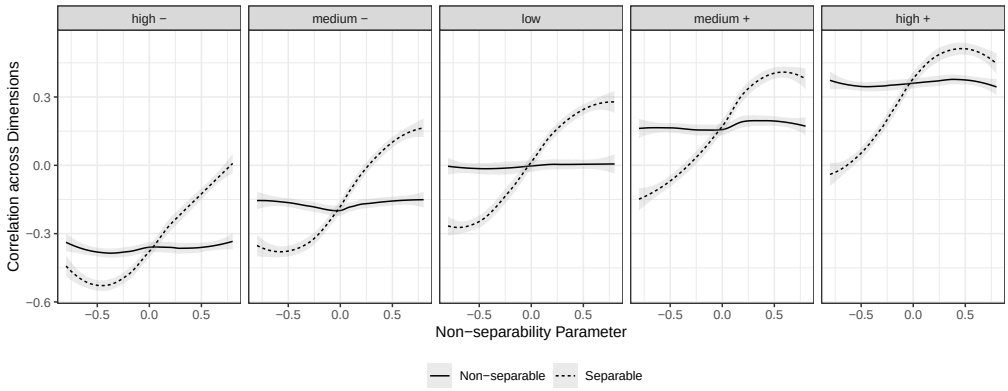



(c) Across correlation of true ideal points.

**Figure 5.** Estimated $\gamma$.

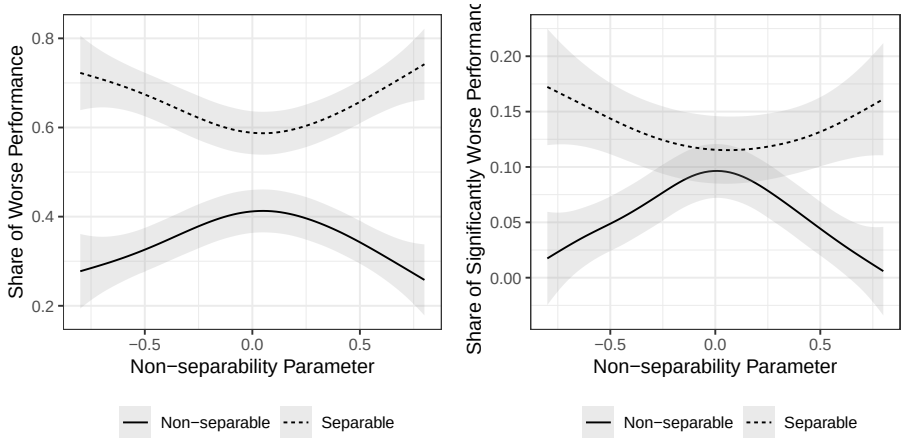



(a) Across salience of primary dimension.




(b) Across correlation of true ideal points.

**Figure 6.** Correlation of estimated ideal points across dimensions.




(a) Share of worse performance.

(b) Share of significantly worse performance.

**Figure 7.** LOO model comparison.

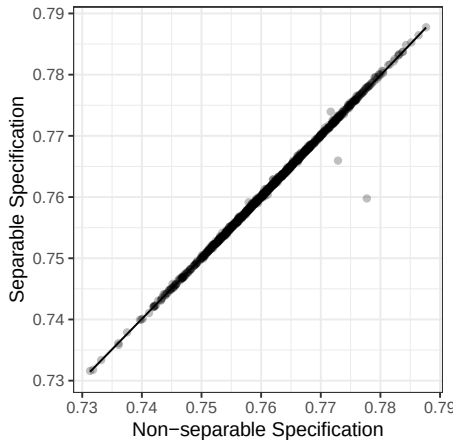



**Figure 8.** Predictive accuracy across specifications.

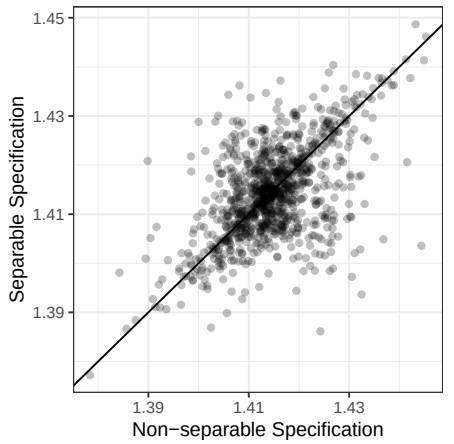



**Figure 9.** Proposal complexity across specifications.

## 3 Application: US Senate

We use data made available by Lewis *et al.* (2020) for this application. In each two-year Congress, around 102 Senators (min.: 97, max.: 107) cast votes on an average of 529 roll calls (sd = 217). Throughout, we exclude unanimous and near-unanimous votes (where less than 5% of all Senators cast a yea- or nay-vote) as well as votes or legislators with less than ten observations overall.

The parametrization of the ideal points along the two dimensions (inter-party and intra-party) is defined by assigning Democrats a group-level indicator value of -1 and Republicans a value of 1 on the first dimension (independent members are set to 0). The second dimension is defined by assigning Senators from coastal states (CA, OR, WA, ME, NH, MA, NY, NJ and CT) a value of -1 and those from former states of the Confederacy a value of 1 (SC, MS, FL, AL, GA, LA, TX, VA, AR, TN and NC). All others are assigned a value of 0. Combined with the constraint on the coefficients of these variables to be >0, this places Democrats at the lower end and Republicans at the upper end of the latent inter-party dimension, and Senators from coastal states at the lower end and senators from states of the former Confederacy at the upper end of the intra-party dimension.

To aid the identification of these dimensions across time, we leverage the fact that senators are often incumbents. Assuming that the meaning of dimensions and the ideological distribution of senators remain somewhat constant across time (e.g., Poole 2007), we incorporate a legislator-specific shift from the group mean observed in each prior Congress in the ideal point predictors of incumbent legislators (i.e. more left-wing Democrats relative to the overall mean of Democrats in the last Congress carry over this residual to the next Congress). This stabilizes our identification strategy (Bafumi *et al.* 2005, p. 178; Gelman and Hill 2006, p. 318).

Figure 10 shows the correlation between estimated ideal points from both specifications. While estimated ideal points on the inter-party dimension are very similar across the two specifications, this is less strongly the case for the intra-party dimension in the period just after the Civil Rights Era, where the correlation drops to values between .6 and .8.

The resulting estimates regarding the ideal points' parametrization are shown in Figure 11 and Figure 12. The $R^2$ statistic presented in Figure 11 shows the extent to which the estimated ideal points are determined by their systematic component (Gelman *et al.* 2019). In most cases, the non-separable specification attributes more of the estimated ideal points to the systematic rather than the stochastic component of their parametrization compared to the separable specification. The estimated coefficients relating covariates to estimated ideal points are shown in Figure 12. In many cases, the separable specification attributes a larger impact to $\delta_{\theta_{d,t-1}}$ rather than the non-separable specification, which conversely estimates a larger impact due to the group-level predictors.

Figure 13 shows the Congress-over-Congress correlation of mean estimated ideal points of Senators by model specification (non-separable and separable). In both specifications, the ideal points of Senators on the inter-party dimension are very robust from one Congress to the next, with a mean correlation of .95 (sd = .03). We interpret this as an indication that the substantive meaning of this dimension is very consistent across time. This is less strongly the case for the intra-party dimension shown in the lower row of Figure 13 with a mean correlation of .84 (sd = .13). The largest shifts along this dimension occur in the 103rd Congress (with a correlation of .52 in the non-separable specification, and a correlation of .44 in the separable specification) and the 111th Congress (with a correlation of .39 in the non-separable specification, and a correlation of .43 in the separable specification). Overall, we take this as an indication that the meaning of the intra-party dimension capturing the remaining variation in roll call votes not attributable to partisan lines varies more strongly, especially in more recent times in which the inter-party dimension has become the dominant dimension in the US Senate (as visible in the main text).

Estimated ideal points for legislators in the US Senate are shown in Figure 14 (non-separable specification) and Figure 15 (separable specification). The solid line indicates the linear correlation between ideal points across dimensions. Focusing on the period after the 89th Congress (charac-

terized by both high levels of non-separability and two similarly salient dimensions), it is evident that the separable specification integrates non-separability by estimating correlated ideal points and that this is much less the case for the non-separable specification.

The difference between estimated ideal points from both specifications is shown in Figure 16 and in Figure 17. In Figure 16, arrows start at the mean estimated ideal point from the separable specification and end at the estimate based on the non-separable specification. This highlights how non-separability can influence estimates of preferences of legislators, by shifting these either mainly along a single dimension (as, e.g., in the 101st and 102nd Congress) or by shifting them across dimensions (as in the Congresses 93–97). In Figure 17, the mean Euclidean distance per Senator between estimated ideal points from either specification is shown relative to the mean estimated non-separability parameter in each Congress. In general, higher levels of non-separability are associated with larger changes in ideal points between the two specifications.

Model comparison based on the LOO information criteria is shown in Figure 18 (only the worse-performing specification per Congress is shown). Negative values imply that a specification performs worse in a pairwise comparison and confidence intervals are calculated based on 2 standard errors of the difference in means (Vehtari, Gelman, and Gabry 2017). The main takeaway of this figure is that neither specification consistently outperforms the other one based on this metric. This is due to the fact that both specifications estimate alternative sets of latent parameters to explain the same observed variance. This results in nearly identical values of information criteria for both models, so that the differences between the two are minute: the absolute differences presented in Figure 18 correspond to relative differences of less than .1% compared to the average information criteria of both specifications per Congress.

A comparison of the predictive accuracy of either specification is shown in Figure 19. Similar to the Monte Carlo results, both specifications perform equally well relative to their fit to the data. As all estimated parameters are latent and we impose only weak constraints for identification, either set of parameters can fit the data similarly well.

Finally, we present the mean complexity of proposals per Congress and specification in Figure 20. Lower values (towards 1) imply that proposals generally only discriminate on one dimension, while higher values (towards 2) imply that they discriminate equally on both dimensions. Overall, the non-separable specification estimates dimensional discrimination structures of proposals that tend to discriminate more uniquely on either of the two dimensions compared to the separable specification. This is shown by the figure because most points lie above the identity line, i.e., proposal complexity is higher in the separable rather than the non-separable specification.

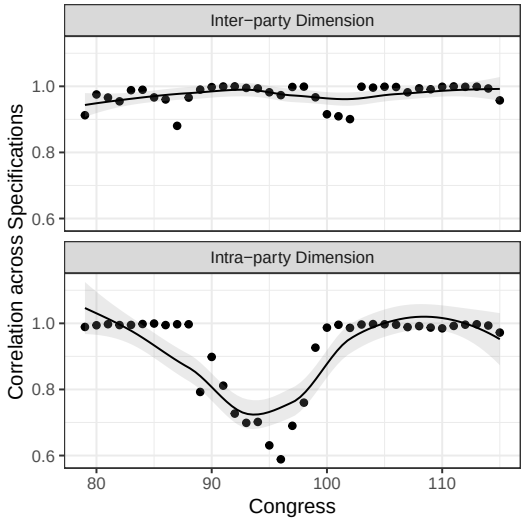



**Figure 10.** Correlation of ideal points across specifications and dimensions.

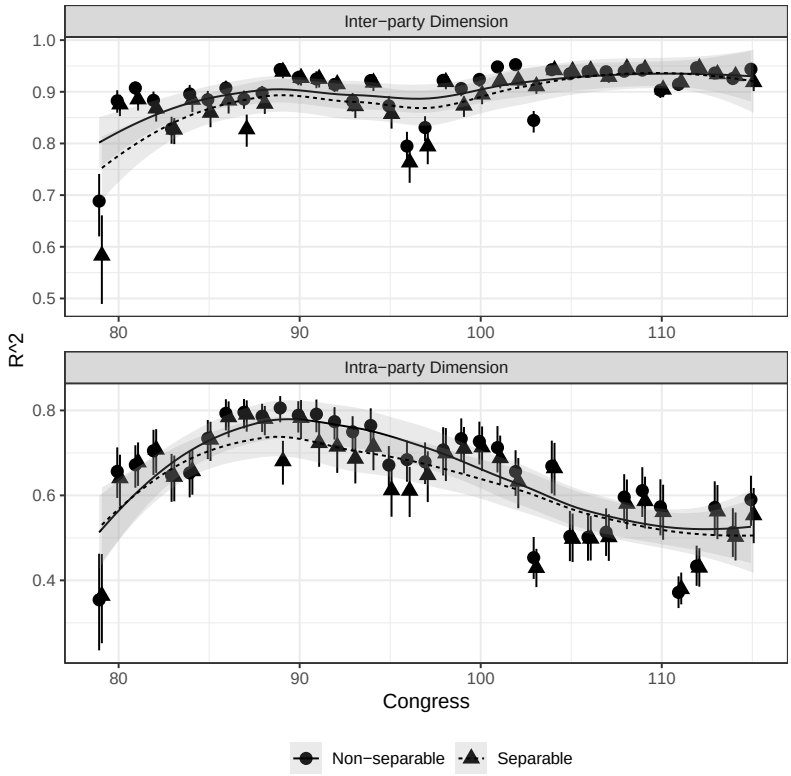



**Figure 11.** Estimated $R^2$ of parametrization of $\boldsymbol{\theta}$.

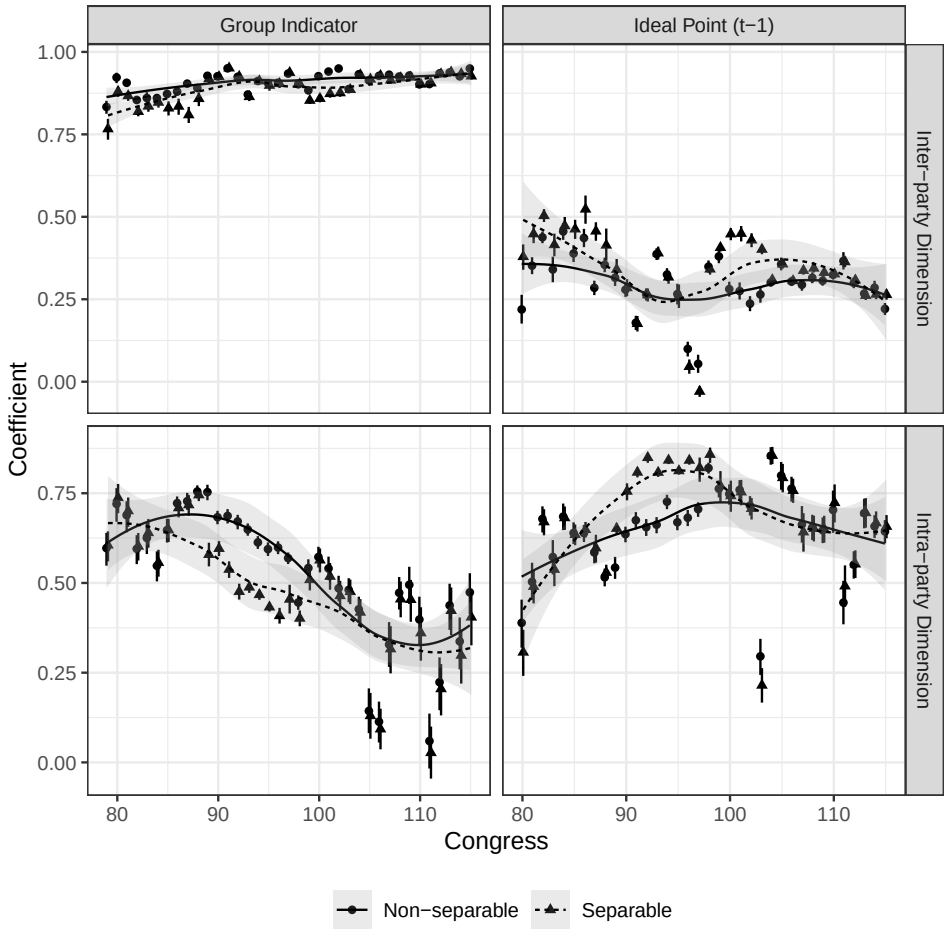



**Figure 12.** Estimated coefficients of parametrization of $\boldsymbol{\theta}$.

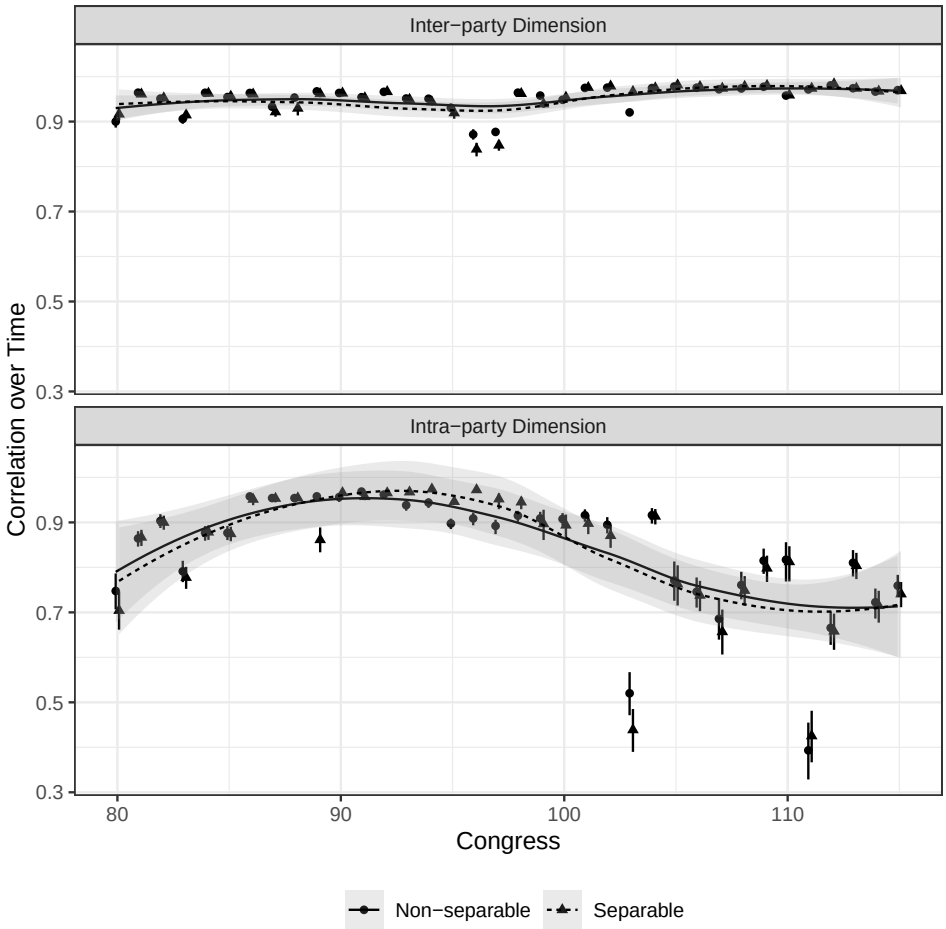



**Figure 13.** Correlation of $\boldsymbol{\theta}$ across time and dimensions.

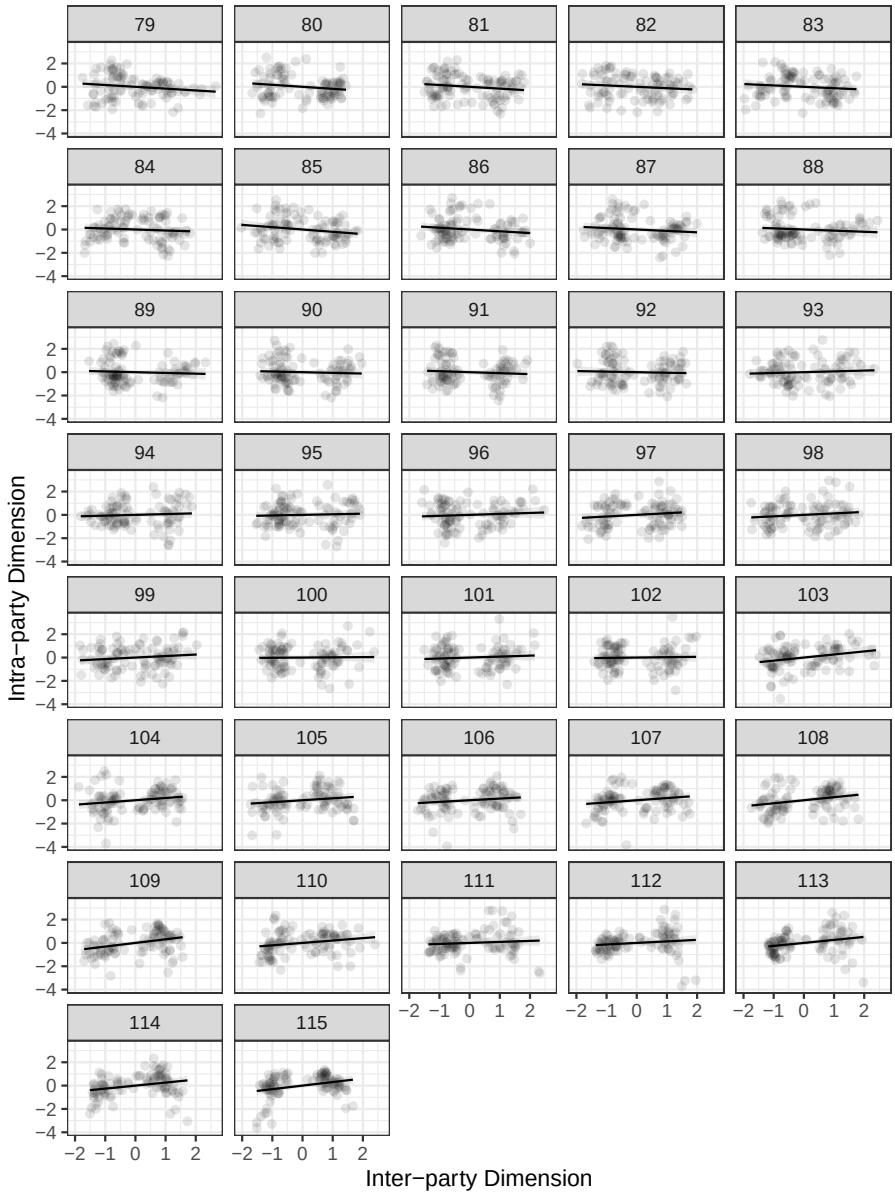



**Figure 14.** Estimated $\boldsymbol{\theta}$ from non-separable specification.

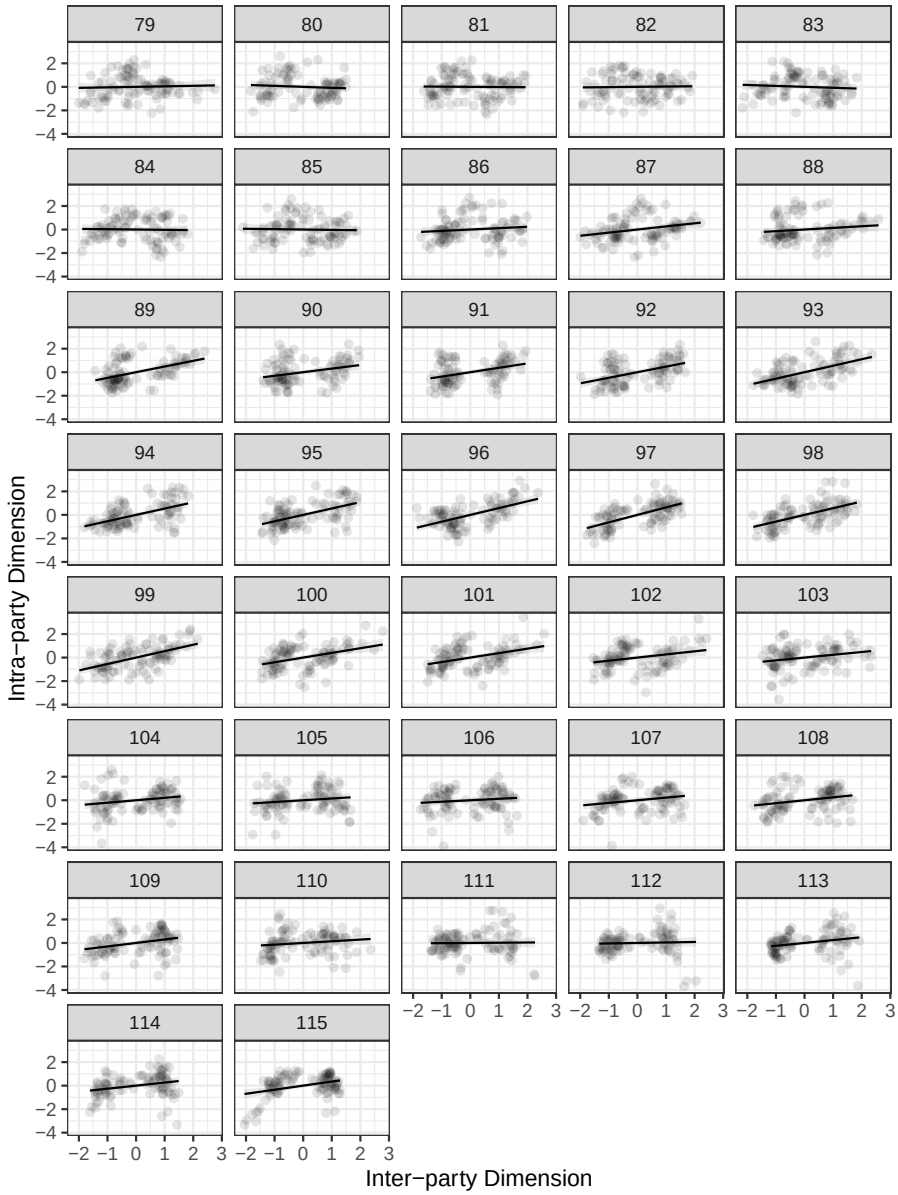



**Figure 15.** Estimated $\boldsymbol{\theta}$ from separable specification.

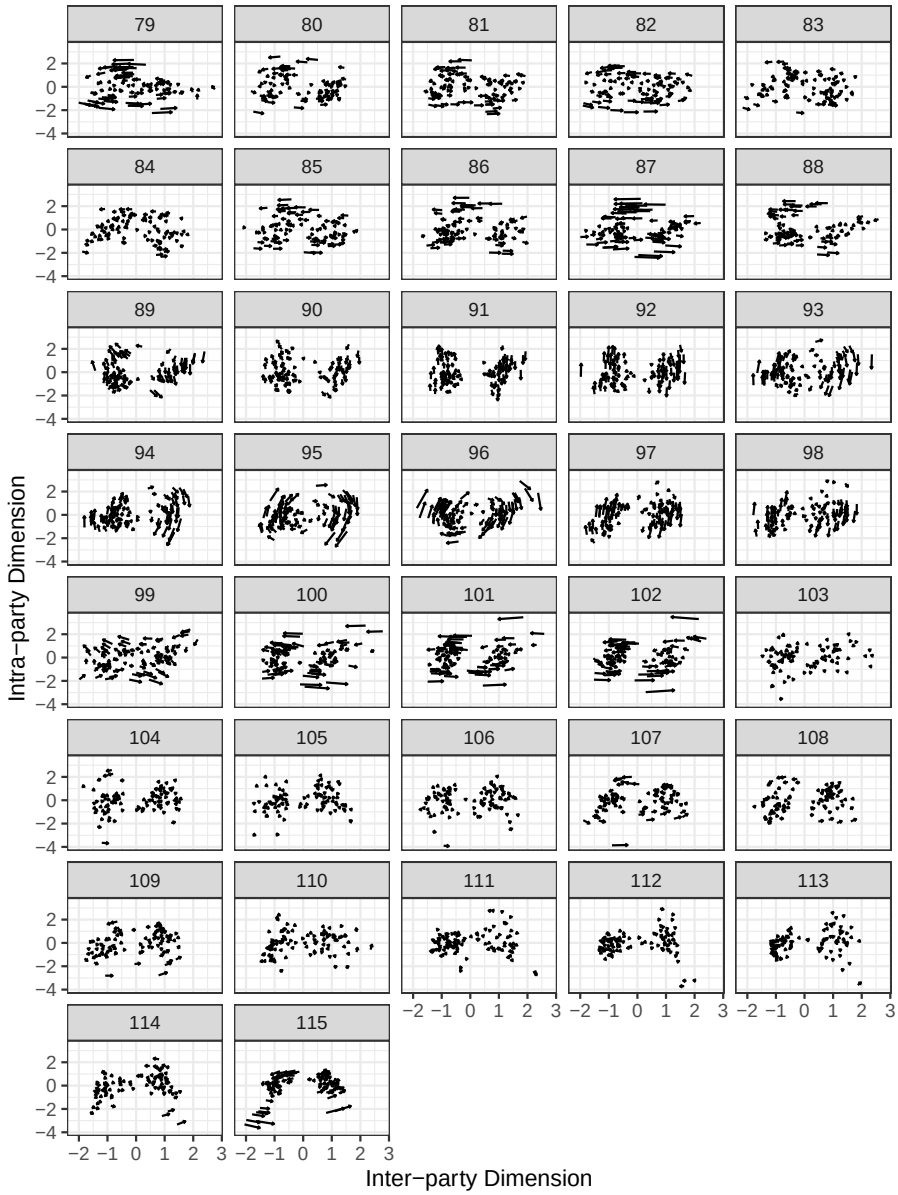



**Figure 16.** Difference of estimated ideal points $\boldsymbol{\theta}$ between separable and non-separable specification.

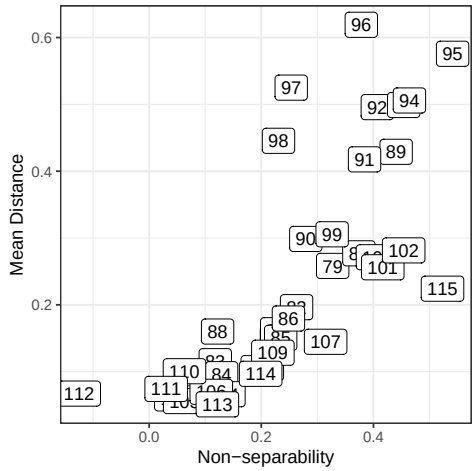



**Figure 17.** Non-separability and difference of estimated ideal points $\boldsymbol{\theta}$ between separable and non-separable specification.

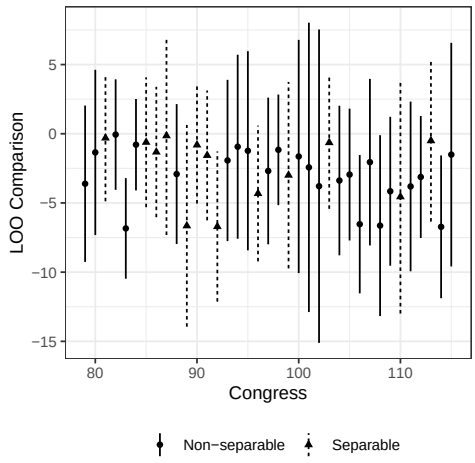



**Figure 18.** Model comparison via LOO.

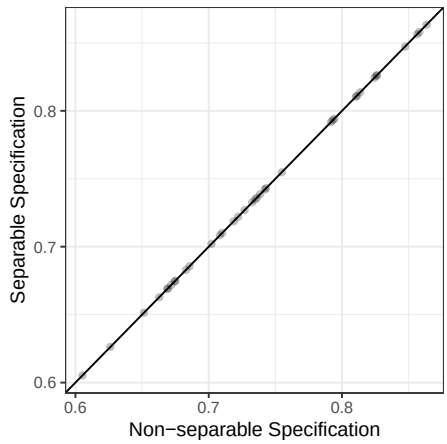



**Figure 19.** Predictive accuracy across specifications.

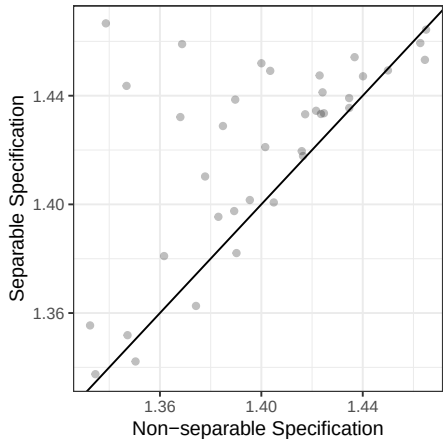



**Figure 20.** Proposal complexity across specifications.

## 4 Application: European Parliament

We draw on data provided by Hix, Noury, and Roland (2006) in this application. We sample 500 roll call votes per five-year session to reduce the computational load during estimation. Per session, an average of 626 legislators cast votes (ranging from 522 in the first session to 719 in the fourth session; sd = 79). Throughout, we exclude unanimous and near-unanimous votes (where less than 5% of all legislators cast a yea- or nay-vote) as well as votes or legislators with less than ten observations.

The two dimensions are defined by the parametrization of the ideal points of legislators. The first left/right dimension is defined by assigning members of the British Conservatives and allies, the Christian Democrats and Conservatives, as well as the French Gaullists and allies a group-level indicator value of 1 (higher values imply more right-wing positions) and those of the Radical left and Greens a value of -1. All others are set to 0. For the secondary EU dimension, members of the Socialists and British Conservatives and allies are set to -1, whereas those of the French Gaullists and allies, Radical left, Anti-Europeans, Greens, and Radical right are set to 1. All others are set to 0. The coefficients of these covariates are constrained to be positive, thereby excluding the possibility of rotation along the dimensions. Additionally, we include $\delta_{\theta_{d,t-1}}$ as discussed in Section 3 of the Appendix.

Figure 21 displays the resulting estimates for dimensional salience in the weight matrix $\mathbf{A}$. These estimates are similar across specifications in the first three sessions and somewhat different in the remaining two sessions characterized by high levels of non-separability. Despite this, the correlation of ideal points across specifications is remarkably high in all sessions (Figure 22), as is their correlation over time similar across specifications (Figure 23). Similarly, the extent to which $\boldsymbol{\theta}$ is determined by the systematic component is similar across specifications (Figure 25) as are the estimated coefficients relating covariates to ideal points (Figure 26). The sole difference between the two specifications is visible in the extent to which ideal points are correlated across dimensions in Figure 24 for the fourth and fifth session. Still, these differences regarding estimated ideal points are less pronounced than in the US application which is presumably due to the fact that there is more data available in the EP than in the US Senate. Finally, Figure 27 (non-separable specification) and Figure 28 (separable specification) show the resulting ideal points of legislators from the five sessions in the EP.

The difference between estimated ideal points from both specifications is shown in Figure 29 and in Figure 30. Overall, changes in estimated ideal points are smaller than in the US application. But similar to the US application, changes are larger as non-separability increases as shown in Figure 30.

Model comparison via the LOO information criteria is shown in Figure 31. The pairwise comparison between the two specifications suggests little significant difference between the two specifications.

A comparison of the predictive accuracy of either specification is shown in Figure 32, while the mean complexity of proposals per session and specification in Figure 33. The results are in line with those from both the Monte Carlo simulations and the US Senate: predictive accuracy is equal across specifications, while the non-separable specifications estimates a less complex dimensional discrimination structure for proposals.

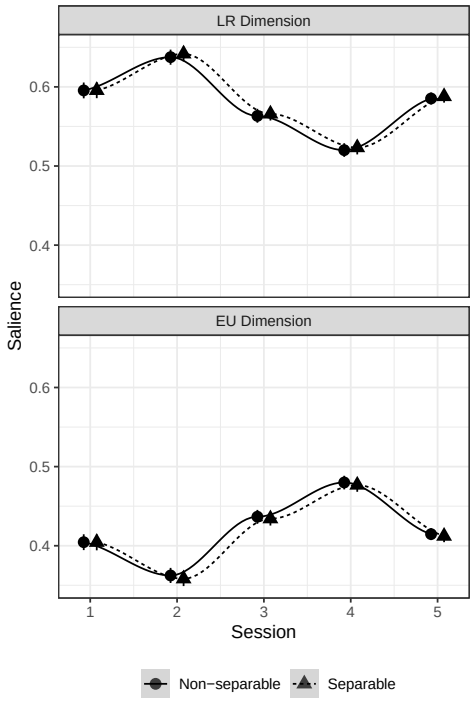



**Figure 21.** Estimated dimensional salience parameters across sessions and dimensions.

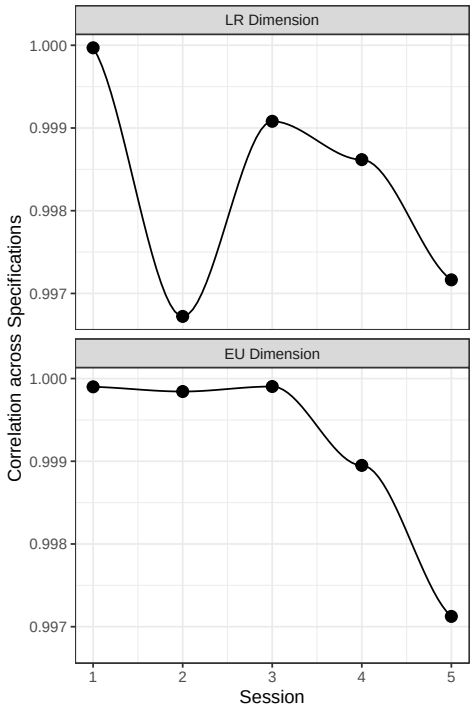



**Figure 22.** Correlation of $\boldsymbol{\theta}$ across specifications.

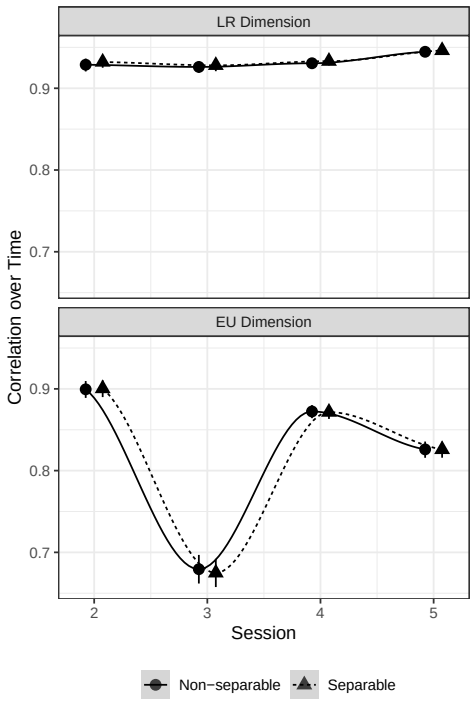



**Figure 23.** Correlation of $\boldsymbol{\theta}$ across time.

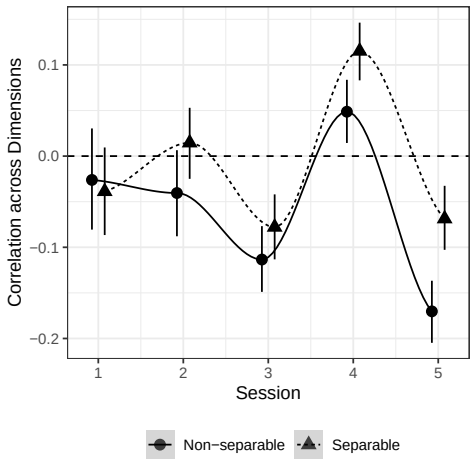



**Figure 24.** Correlation of $\boldsymbol{\theta}$ across dimensions.

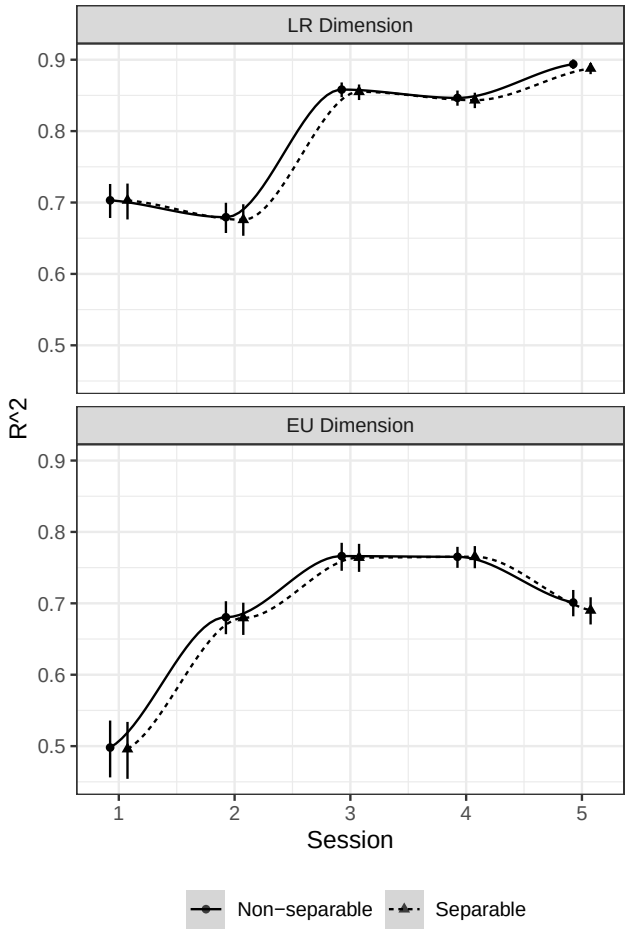



**Figure 25.** Estimated $R^2$ in EP sessions.

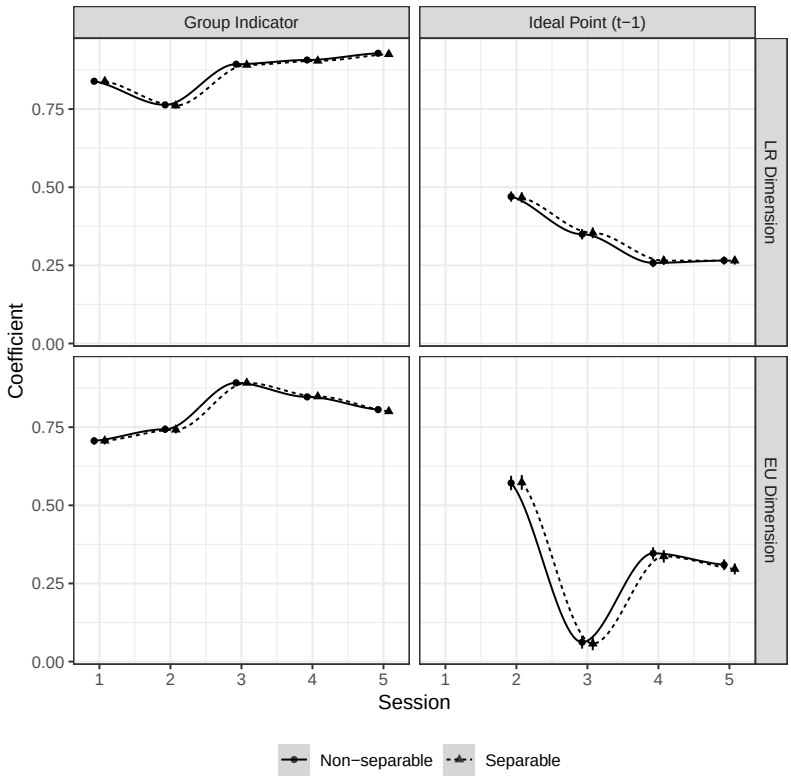



**Figure 26.** Estimated coefficients of parametrization of $\boldsymbol{\theta}$ in EP sessions.

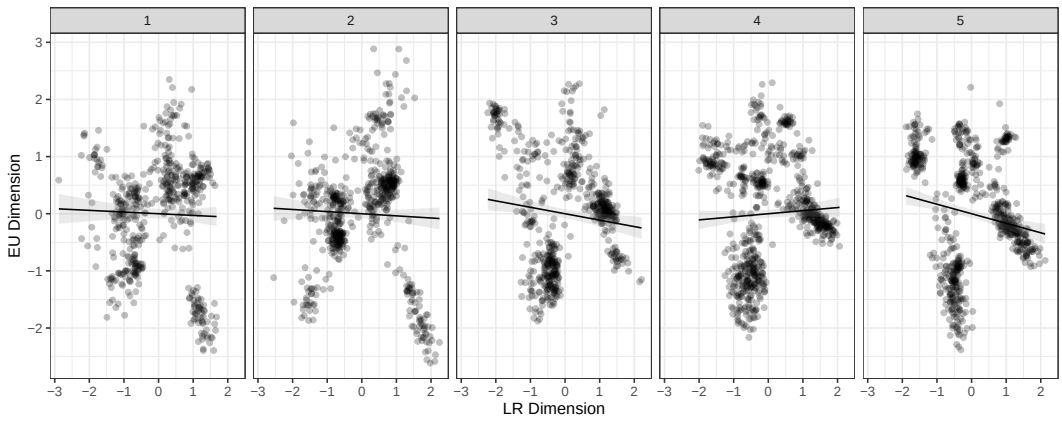



**Figure 27.** Estimated $\boldsymbol{\theta}$ from non-separable specification.

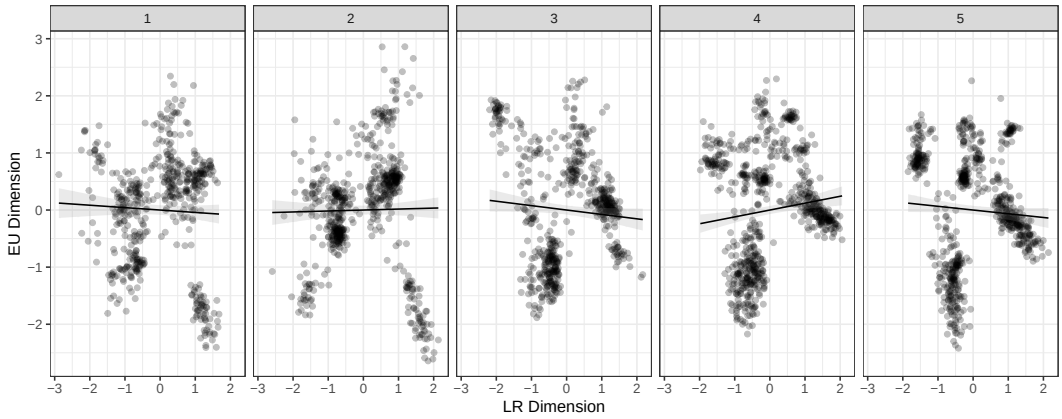



**Figure 28.** Estimated $\boldsymbol{\theta}$ from separable specification.

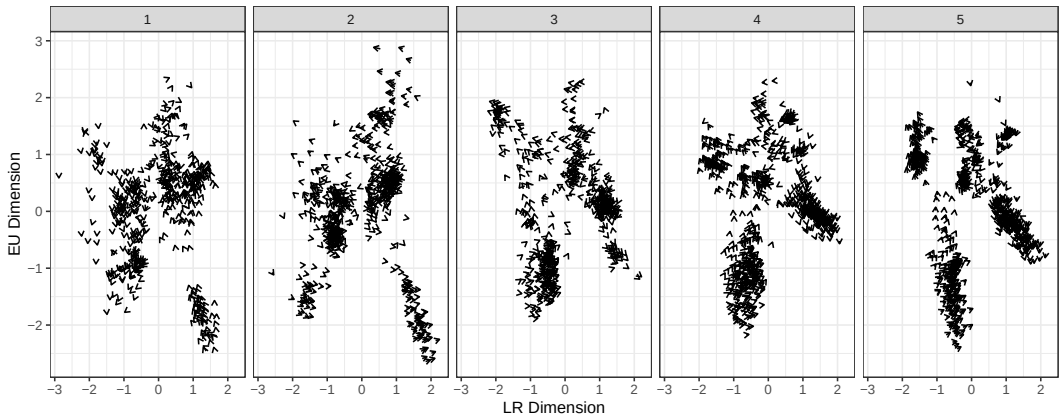



**Figure 29.** Difference of estimated ideal points $\boldsymbol{\theta}$ between separable and non-separable specification.

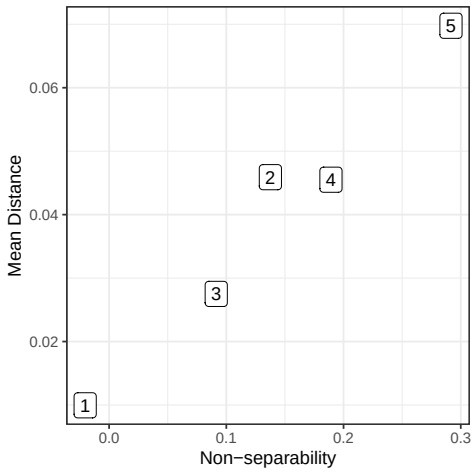



**Figure 30.** Non-separability and difference of estimated ideal points $\boldsymbol{\theta}$ between separable and non-separable specification.

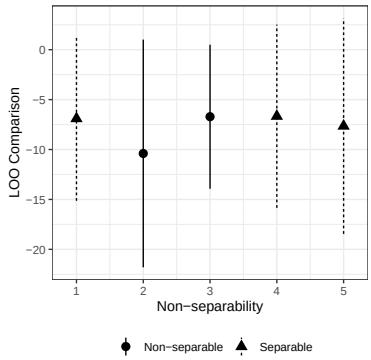



**Figure 31.** Model comparison via LOO.

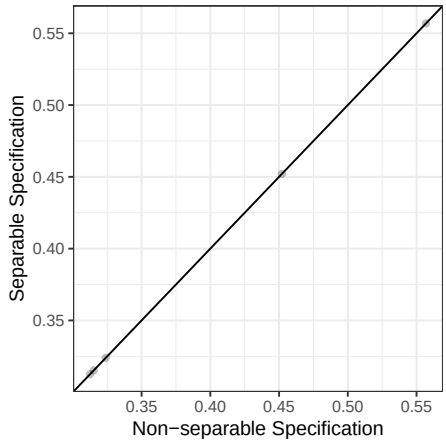



**Figure 32.** Predictive accuracy across specifications.

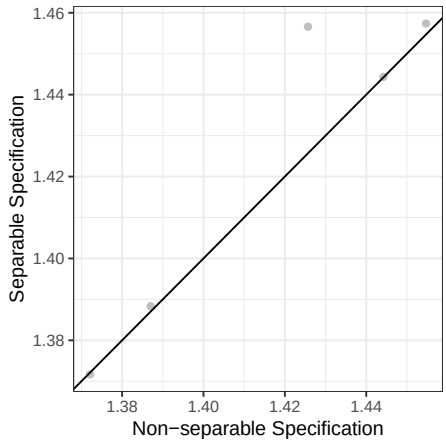



**Figure 33.** Proposal complexity across specifications.