Collected Notes on Aldrich-Mckelevey Scaling

Phil Swatton

Contents

1 Introduction

Aldrich-McKelvey scaling is a method designed to correct differential item functioning in survey respondent placements of external stimuli along ordered rating scales, such as the ideological positions of political parties (Aldrich and McKelvey 1977). In doing so, it allows for recovery of corrected stimuli positions and by extension respondent positions or ideal points along the same scales. Differential item functioning can be understood as survey respondents perceiving or understanding the scale in fundamentally different ways - thus leading to differing placements of the stimuli. Aldrich-McKelvey scaling has sometimes also been discussed in the context of rationalisation bias, where respondents have a tendency to shift the placements of parties they dislike away from themselves and the placements of parties they like towards themselves. However, it is only in the form of a much more recent adaptation that rationalisation bias has been explicitly included in the model, although Aldrich-McKelvey scaling remains robust in the case of estimating stimuli positions and reasonably robust in the case of estimating respondent positions (see Bølstad 2020).

This more recent adaptation was itself based on an adaptation of the Aldrich-Mckelvey scaling method in the form of Bayesian Aldrich-McKelvey scaling (Hare et al. 2015). This has also been developed in an alternative direction, with Ordered Bayesian Aldrich-McKelvey scaling being first proposed at the 2019 MPSA annual meeting (Cikanek, McAlister, and Shin 2019). In a different, older vein, the method was also generalised to a multidimensional case - albeit one in which the ability to scale respondents and stimuli simultaneously was lost (Keith T. Poole 1998). Although not well-known outside political science (and perhaps not even within), application-specific methods such as Aldrich-McKelvey scaling represented in the estimation of Gary King et al the most satisfactory approach to correction for differential item functioning (2004). The method is unsurprisingly particularly popular in work exploring the spatial model of political competition. Results from Aldrich-McKelvey scaling have been used to generate estimates of voter information (Palfrey and Poole 1987), in scaling elite positions (Saiegh 2009), in testing Downsian theories of voter preference (Hollibaugh,

Rothenberg, and Rulison 2013), in placing voters and parties from different countries on the same scale (Lo, Proksch, and Gschwend 2014), and in assessing the role of valence in voter utility functions (Gouret 2021).

What makes Aldrich-Mckelvey scaling and its derivatives unique as compared to other more generic methods such as factor analysis, item response theory, or multidimensional scaling is the fact that it scales the stimuli at *the same level of input* in which they exist. In other words, where other scaling methods necessarily reduce the dimensionality of the data they take as input, Aldrich-Mckelvey scaling retains the same dimensionality for both stimuli and respondents. What it instead does is assign variation in stimuli placements to a set of parameters broadly capturing the process of differential item functioning. This is not to say that models allowing the recovery of higher dimensions are not useful or theoretically important. Rather, this unique feature of Aldrich-McKelevey scaling is useful both for practical purposes and for better exploration of respondent variation in lower dimensionalities. While not well-known outside the world of political science, this may in fact give rise to more use in other social sciences, such as psychology or sociology, where the placement of external stimuli along some continuum may in fact be useful at a particular level of analysis.

Despite the method's usefulness, its several adaptations, and its popularity in certain parts of the political science community, it remains surprisingly under-explored. Widely popular methods such as ordinary least squares are well-understood, with several important results and findings regarding best practice for estimation, identification conditions, and performance under the violation of assumptions. My purpose in this collection of notes is to present several small, key results in these areas, with further comment on differences between Aldrich-Mckelvey scaling and Bayesian Aldrich-Mckelvey scaling. The first note discusses the possibility of estimating the Aldrich-Mckelvey solution through QR decomposition, thus allowing for greater numerical accuracy and better respondent retention. This estimation strategy in turn gives rise to the second note, which shows that the Aldrich-Mckelvey model is not identified when the number of stimuli is below 3. The third note challenges a key assumption in the Aldrich-Mckelvey literature, which is that it is especially robust to heteroskedasticity as compared to mean-centering. As we know from the forecasting aggregation literature, taking the mean is also robust to heteroskedasticity. There remain however good reasons for favouring the Aldrich-McKelvey scaling method seperate from performance assessments. The fourth and final note discusses the fact that the Bayesian Aldrich-Mckelvey parameters as set up in the model necessarily have a different interpretation to the original Aldrich-Mckelvey model, setting asside conceptual differences in the estimation approaches.

2   Aldrich-Mckelvey Scaling

Before proceeding to the first note, I provide a brief sketch of Aldrich-McKelvey scaling and its traditional estimation via a least-squares solution. As above Aldrich-McKelvey scaling is a method for correcting differential item functioning in the placement of external stimuli along an ordered rating scale. The method does not *require* that the data be ordinal (and in fact treats them as continuous), but in practice the data available to researchers are always ordinal. Let n be the number of respondents indexed by i and J be the number of external stimuli indexed by j. We then place the J placements of stimuli by respondent i in a matrix with a column of 1s:

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & \boldsymbol{X}_{i1} \\ 1 & \boldsymbol{X}_{i2} \\ \vdots & \vdots \\ 1 & \boldsymbol{X}_{iJ} \end{bmatrix} \tag{1}$$

Aldrich-McKelvey write a model of the form

$$c_i + w_i X_{ij} = Y_j + u_{ij} \tag{2}$$

where c_i is the intercept or shift parameter for respondent i and w_i is the weight or scaling parameter. I retain the use of 'intercept' and 'weight' respectively in the remainder of this paper. In the Aldrich-McKelvey model, these parameters govern the differential item functioning process. $\boldsymbol{Y_j}$ represents the true position of the jth stimulus, while u_{ij} is an error term, for which Gauss-Markov assumptions apply. To obtain the

solution to the above model, set:

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{X}_i'(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' \tag{3}$$

Aldrich and McKelvey prove that the that the Aldrich-McKelvey estimator $\hat{\boldsymbol{Y}}$ of the 'true' stimuli positions $\boldsymbol{Y}$ is the eigenvector of

$$\boldsymbol{A} - n\boldsymbol{I} \tag{4}$$

corresponding to the highest negative eigenvalue of the above (i.e. the negative eigenvalue closest to 0). With the Aldrich-McKelvey scaling method briefly established (see Aldrich and McKelvey 1977 for the full proof), I now turn to the four notes that compose this paper.

3   Estimation by QR Decomposition

A fact well-known among statistical software developers but much less well-known in the wider social science community is that the ordinary least squares model (OLS) is not estimated through the well-known matrix formula $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}'\boldsymbol{Y}$. Instead, OLS is typically estimated through some form of *matrix decomposition -* usually *QR decomposition*. This is due to the fact that floating-point computations contain some inherent inaccuracy (see Goldberg 1991), thus in turn rendering matrix inversion to be particularly numerically unstable. R's *lm* function in particular represents a popularly-used function that estimates OLS parameters through QR decomposition. It follows that Aldrich-McKelvey scaling may also benefit from an alternative approach to estimation. In this note, I show how QR decomposition can also be used for the purpose of estimating the Aldrich-McKelvey scaling results.

First, a brief examination of (3) shows that Aldrich-McKelvey scaling relies on the calculation of

$$\boldsymbol{A}_i = \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' \tag{5}$$

for all n units. In practice, for some respondents $(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$ may not be invertible. When this is the case, the information provided by some respondents is lost and we use $< n$ respondents in calculating the stimuli positions. However, by using QR decomposition, it becomes possible not only to more accurately calculate $\boldsymbol{A}_i$, but also to retain a greater number of respondents in the overall estimation. While atypical in presenting the QR decomposition, I keep the i index throughout to better relate the decomposition to calculating (5) for each respondent. The QR decomposition of a matrix is given by

$$\boldsymbol{X}_i = \boldsymbol{Q}_i\boldsymbol{R}_i \tag{6}$$

where $\boldsymbol{R}_i$ is an upper triangular matrix and $\boldsymbol{Q}_i$ is of the same dimensions of $\boldsymbol{X}_i$ and its columns are orthonormal, meaning that

$$\boldsymbol{Q}_i'\boldsymbol{Q}_i = \boldsymbol{I} \tag{7}$$

Substituting the QR decomposition from (6) into the (5) we obtain

$$\boldsymbol{A}_i = \boldsymbol{Q}_i\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{Q}_i'\boldsymbol{Q}_i\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'\boldsymbol{Q}_i' \tag{8}$$

recalling (7) this simplifies to :

$$\boldsymbol{A}_i = \boldsymbol{Q}_i\boldsymbol{R}_i(\boldsymbol{R}_i'\boldsymbol{R}_i)^{-1}\boldsymbol{R}_i'\boldsymbol{Q}_i' \tag{9}$$

The rules of algebraically manipulating matrices dictate that

$$(\boldsymbol{R}_i'\boldsymbol{R}_i)^{-1} = \boldsymbol{R}_i^{-1}(\boldsymbol{R}_i')^{-1} \tag{10}$$

and we can thus substitute (10) into (9) to obtain:

$$\boldsymbol{A}_i = \boldsymbol{Q}_i\boldsymbol{R}_i\boldsymbol{R}_i^{-1}(\boldsymbol{R}_i')^{-1}\boldsymbol{R}_i'\boldsymbol{Q}_i \tag{11}$$

which by the definition of an inverse and an identity matrix simplifies to

$$\boldsymbol{A}_i = \boldsymbol{Q}_i \boldsymbol{Q}_i' \tag{12}$$

thus drastically simplifying (3). This simplification avoids the need to invert

$$\boldsymbol{X}_i' \boldsymbol{X}_i \tag{13}$$

which as above not only is numerically unstable in practice but also results in the unnecessary loss of some respondents from the calculation of placements for the stimuli. This solution instead reduces the problem of estimating (5) to a problem of estimating the QR decomposition for each respondent, rather than performing a direct inversion for each respondent. However, this result for the purpose of estimation also aids us in better understanding the identification conditions for Aldrich-McKelvey scaling.

4 Identification

An important consideration for many types of latent variable model or scaling method is that to some degree model identification is driven by the need to put information in to get information out. Factor analysis - both exploratory and confirmatory - has strict rules on the number of parameters, variances, and covariances being estimated relative to the number of variances and covariances being inputed. Principle components analysis estimates only as many principle components as input vectors. The question then follows - what are the identification constraints for Aldrich-McKelvey scaling? While Aldrich and McKelvey place some scale constraints (e.g. mean 0) (1977), this is to prevent indeterminacy in the solution for any value of J. The question here is what is the minimum J for the model to be identified - i.e. for a unique solution to exist for some given data. Using QR decomposition, it is possible to show that no solution exists once $J < 2$. The result that follows begins by proving this in the case of $J = 2$, before discussing the case of $J = 1$.

4.1 No Unique Solution when $J = 2$

We begin by recalling the QR-decomposition based approach to calculating $\boldsymbol{A}$ for the Aldrich-McKelvey solution: Recall from the new approach to estimating the Aldrich-McKelvey solution via the QR decomposition that

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{A}_i = \sum_{i=1}^{n} \boldsymbol{Q}_i \boldsymbol{Q}_i' \tag{14}$$

When $J = 2$, each $\boldsymbol{X}_i$ matrix and thus each $\boldsymbol{Q}_i$ matrix will be square. When the QR decomposition is performed on a square matrix, the Q matrix will be an orthogonal matrix meaning that

$$\boldsymbol{Q}_i' \boldsymbol{Q}_i = \boldsymbol{Q}_i \boldsymbol{Q}_i' = \boldsymbol{I} \Rightarrow \boldsymbol{Q}_i' = \boldsymbol{Q}_i^{-1} \tag{15}$$

it follows that in this special case:

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{Q}_i \boldsymbol{Q}_i' = \sum_{i=1}^{n} \boldsymbol{I} = n\boldsymbol{I} \tag{16}$$

If we insert (16) into (4) we obtain

$$n\boldsymbol{I} - n\boldsymbol{I} \tag{17}$$

meaning that when $J = 2$ the eigendecomposition will be performed on a 2×2 $\boldsymbol{0}$ matrix.

When an eigendecomposition is performed on any $\boldsymbol{0}$ matrix, all eigenvalues will be 0 and thus the number of corresponding eigenvectors of the correct proportion will be infinite. In other words, there is no solution. This follows first because there is no negative eigenvalue (recall that the Aldrich-McKelvey solution selects based on the highest negative eigenvalue); and second even ignoring this the range of corresponding eigenvectors to a 0 eigenvalue is infinte. There is therefore no unique solution, and thus the scaling method is not identified when $J = 2$. Researchers should be cautious in this case: the fact that the result has an infinite range of

solutions does not mean that software will not present a solution. Running base R's **eigen** function on a 2×2 **0** matrix produces a matrix with two eigenvectors containing the arbitrary values of -1 and 1:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{18}$$

Similarly and more importantly, the **aldmck** function from the **basicspace** R package will produce the values of 0 and 1 for its stimuli placement estimates despite correctly returning a full set of 0s for the extracted eigenvalues (Keith T. Poole et al. 2016). Researchers utilising this package (and given it is to my knowledge the only package making the Aldrich-McKelvey scaling method widely accessible they are likely to be) should take care in interpreting its results - any rescaling performed when $J < 3$ will necessarily be meaningless as they stem from a solution which in practice is not identified and thus the results themselves could change arbitrarily.

4.2 No Unique Solution when $J = 1$

Until this point, in this note I have focussed on the case where $J = 2$ while making claims regarding the broader case of $J < 3$. I now turn to the case of $J = 1$. While intuitively it makes sense for a model to be unidentified when $J = 1$ if it is unidentified when $J = 2$, it is nonetheless worth establishing this beyond the realm of intuition. First, estimation of (12) cannot occur when $J = 1$ because the QR decomposition only exists when the number of rows is greater than or equal to the number of columns. It follows that the calculating $\boldsymbol{A}$ must necessarily fall back to (3) (insofar as other methods of calculating $\boldsymbol{A}$ have not yet been discovered.

However, in the case of $J = 1$ calculation of each $\boldsymbol{X}_i' \boldsymbol{X}_i$ takes on a particular behaviour that prevents its own inversion. To show this, it is worth demonstrating how $\boldsymbol{X}_i' \boldsymbol{X}_i$ is calculated if its two component values are not known when we have two columns and one row. We begin by setting the two values of $\boldsymbol{X}_i$ to a and b:

$$\boldsymbol{X}_i = \begin{bmatrix} a & b \end{bmatrix} \tag{19}$$

Taking the inner product of (19) will take the form

$$\boldsymbol{X}_i' \boldsymbol{X}_i = \begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} \tag{20}$$

which necessitates the following multiplications of a and b:

$$\boldsymbol{X}_i' \boldsymbol{X}_i = \begin{bmatrix} a^2 & b*a \\ a*b & b^2 \end{bmatrix} \tag{21}$$

In the case of Aldrich-McKelvey scaling, by the defintion of $\boldsymbol{X}_i$ in (1) we know that $a = 1$. Inserting this into (21) yields the following solution for $\boldsymbol{X}_i' \boldsymbol{X}_i$:

$$\boldsymbol{X}_i' \boldsymbol{X}_i = \begin{bmatrix} 1 & b \\ b & b^2 \end{bmatrix} \tag{22}$$

A clear feature of the $J = 1$ solution for $\boldsymbol{X}_i' \boldsymbol{X}_i$ is that the second row is a simply the first multiplied by b (and similarly, the second column is simply the first multiplied by b). This means that the matrix $\boldsymbol{X}_i' \boldsymbol{X}_i$ is singular and its inverse does not exist. While it is therefore possible to show that a solution does not exist for the Aldrich-McKelvey solution does not exist either via traditional estimation or via the QR decomposition above due to the impossibility of calculating $\boldsymbol{A}$, a more conclusive proof of the type existing for the case of $J = 2$ does not yet exist. It is likely that other matrix decompositions could be used to calculate $\boldsymbol{A}$ - we must rely on the non-existence of a solution when $J = 2$ and a the non-existence of a solution for both approaches to estimation when $J = 1$ to inductively conclude that a solution does not exist when $J = 1$.

5 Central Tendency as an Estimator of Stimulus Position

It is frequently asserted that the advantage of Aldrich-McKelvey scaling relative to simply taking the mean value of a vector of placements is that Aldrich-McKelvey scaling is robust to heteroskedasticity in the error term u_{ij} in (2). This claim originates from a paper by Palfrey and Poole (1987), where they show that Aldrich-McKelvey scaling is robust to heteroskedasticity in the u_{ij} in recovering stimuli placements but has a bias towards unidimensional results that becomes stronger for low-information respondents. After this initial piece of research, the claim that Aldrich-McKelvey scaling should be preferred to taking the mean *because it is robust to heteroskedasticity where taking the mean is not* proliferated (see e.g. Armstrong et al. 2020, 17; Gouret 2021). However, by drawing comparisons with findings from the world of forecasting and forecasting aggregation, it becomes possible to realise that taking the mean is just as robust to heteroskedasticity as Aldrich-McKelvey scaling.

The comparability of forecast aggregation and Aldrich-McKelvey scaling stems from the fact that both represent 'crowd wisdom' approaches to estimating some quantity of interest. In the case of forecast aggregation, our interest is in attempting to quantify the probability of some future event occurring. In the case of Aldrich-McKelvey scaling, our interest is instead in translating the 'common sense' of voters regarding party positions into quantifiable party placements alongside respondent placements. The difference of course is that in the case of forecast aggregation, we eventually learn whether an event occurs or not and in the long term can learn how accurate different aggregators are. By contrast, with Aldrich-McKelvey scaling and its derivatives we never learn at any point in time what the 'true' position of some political party is along some dimension of interest. We are therefore constrained to relying on a combination of assumptions about the data generating process, things we can actually learn about the data generating process without observing 'true' values, and simulations when seeking to understand the behaviour of the Aldrich-McKelvey scaling method.

Of interest to me in this note is the 'Bias-Information-Noise' (BIN) model for assessing forecasts (Satopää et al. 2021). As suggested by the name, this model decompses the results of forecasts into three key components. Importantly, this allows us to assess how different forecast aggregators actually work - do they reduce bias, increase the information recovered, or simply reduce noise? If θ is some estimated value and T is the 'true' value, then the BIN model can be summarised in the following three formulae:

$$
\begin{aligned}
Bias &= E[\theta - T] \\
Information &= cov(\theta, T) \\
Noise &= var(\theta) - var(T)
\end{aligned}
\tag{23}
$$

An important finding from the world of forecast aggregation is that taking the mean is an effective way of reducing *noise*, but not necessarily of increasing *information* or reducing *bias* (Satopää 2017; Satopää et al. 2021). The implication here is that the mean will reduce any form of *noise* but will not remove structural biases. Whether that noise increases or decreases linearly with some variable - such as the amount of information a respondent possesses - is not necessarily important. Taking the mean will still reduce the noise to a single central point. Given that Palfrey and Poole (1987) only assess the violation of the homoskedasticity Gauss-Markov assumption and thus is not a form of *bias* per se, there is no reason to assume that taking the mean should be any less robust to heteroskedasticity in (2). Indeed, this is directly testable. Taking the simulation code demonstrating the robustness of Aldrich-McKelvey scaling in the article for the **basicspace** R package (Keith T. Poole et al. 2016), it is easy to see that the means are infact highly correlated with the Aldrich-McKelvey solution:

```r
# For replicability
set.seed(1234)

# Stimuli positions
J <- 6
Yj <- rnorm(6, mean=0, sd=1)
```

```r
Yj <- (Yj-mean(Yj))/sd(Yj)

# Respondent error
N <- 500
resp_sd <- runif(N, min=0.3, max=0.9)
ui <- matrix(NA, N, J)
for (i in 1:N) ui <- rnorm(N, mean=0, sd=resp_sd)

# Respondent intercepts and weights
wi <- runif(N, min=0, max=1)
ci <- rnorm(N, mean=0, sd=1)

# Respondent placements
Yij <- (rep(1,N) %o% Yj) + ui
Xij <- (1/wi) * (Yij - ci)

# aldmck from basicspace
am <- aldmck(Xij, polarity=1)
stim <- am$stimuli

# Taking the means
means <- apply(Xij, MARGIN=2, FUN=mean)
```

The above code is adapted from Poole et. al. (2016), with changes to the code to make it consistent with the notation I've used in these notes and the addition of a seed to facilitate replicability. Note the heteroskedastic error u_i instead of the previous error term u_{ij} from (2). I have also slightly tidied some steps to simplify the code. In line with past results, the stimuli placements are closely matched to the 'true' stimuli positions with a Pearson's correlation of 1. However, so too are the means with a Pearson's correlation of 1 (both are not exactly 1 but have been rounded to the nearest 3 decimal places, giving a correlation of effectively 1). Indeed, the correlation between the recovered stimuli positions from the Aldrich-McKelvey scaling and simply taking the means is 1 (once again after rounding). Given the similarity between these results and the relative ease of taking the means, why then use Aldrich-McKelvey scaling?

The most straightforward answer to the above question is that it renders the assumptions of our process *transparent*. When we perform Aldrich-McKelvey scaling, there are clear assumptions regarding the data generating process as articulated in the original paper (Aldrich and McKelvey 1977). It is clear that insofar as recovery of the stimuli is concerned, both methods are robust to heteroskedasticity in (2). But when we take the mean, it is no longer clear whether we are assuming a data generating process of the form in (2) with all the corresponding assumptions. Similarly, while we *could* regress the mean placements on the respondent placements to obtain similar parameters, there is less transparency in the process by which we obtain these parameters and the assumptions we make in the process. Where the assumptions of the Aldrich-McKelvey scaling method *are* violated, this creates space for the development of new scaling methods that can be built on this initial attempt.

6 BAM is not AM

By far one of the most popular adaptations of Aldrich-McKelvey scaling is Bayesian Aldrich-McKelvey scaling. The popularity of the Bayesian variant is unsurprising given two key limitations of the original version. First, the only process developed for obtaining error bounds on the Aldrich-McKelvey stimuli placements is bootstrapping. Second and more importantly, Aldrich-McKelvey scaling has strict requirements for the number of placements respondents make - if some respondent places say 4 out of 6 stimuli, the information that respondent possesses is necessarily lost. While the 'blackbox' scaling method allowed the presence of missing data (Keith T. Poole 1998), this method necessarily scales data to a higher dimensionality and does not scale respondents and stimuli simultaneously. Bayesian Aldrich-McKelvey scaling fixes both of these

issues - first by allowing missing data in stimuli placements; and second by using the posterior distributions for

These gains facilitate a number of research applications not previously imaginable for Aldrich-McKelvey scaling. First, where the anchoring vignettes first proposed to serve as a means of establishing the comparability of different stimuli placements across space and time (King et al. 2004) had previously been used in the 'blackbox' scaling method (Bakker et al. 2014); Bayesian Aldrich-McKelvey scaling now allowed this to be performed at the lower level of dimensionality (Jessee 2021). This allows not only for stimuli positions to be compared at lower dimensionalities over time and space, but also for respondent positions along the same scale. None of this necessarily precludes exploring the higher dimensionalities or 'basic space' of ideology - but it does allow these explorations to begin from scales corrected for differential item functioning.

However, a point that has not yet appeared in the literature is the fact that Bayesian Aldrich-McKelvey scaling is built on a slightly different data generating process to the original Aldrich-McKelvey scaling model. Recall that the Aldrich-McKelvey model was of the form

$$c_i + w_i X_{ij} = Y_j + u_{ij} \qquad \text{(2 repeated)}$$

By contrast, the Bayesian-Aldrich McKelvey scaling model is of the form (Hare et al. 2015)

$$a_i + b_i Y_j = X_{ij} + u_{ij} \qquad (24)$$

where a_i is the intercept parameter and b_i is the weight parameter. The notable difference from (2) to (24) is the fact the parameters operate on Y_j instead of X_{ij}. Insofar as I have been able to tell, this is not a delibrate change - the same difference is repeated in describing the original Adlrich-McKelvey model in the *Analysing Spatial Models of Judgement of Choice* book (Armstrong et al. 2020). This is not necessarily straightforwardly important, but it does mean the parameters have differing interpretations. Arguably, the interpretation of the Bayesian Aldrich-McKelvey parameters is more in line with with our interpretation of the model in that they 'shift' and 'stretch' the true stimulus position to the perceived one thus directly capturing the differential item functioning process. By contrast, the Aldrich-McKelvey parameters relate the perceived stimulus to the 'true' stimulus. Insofar as the difference is a mistake it is therefore a happy one. This is worth noting not just for comparisons between the individual-level parameters of the results. This difference between the models has also made it into adaptations of the Bayesian Aldrich-McKelvey scaling method such as Ordered Bayesian Aldrich-McKelvey scaling (Cikanek, McAlister, and Shin 2019) and the Intercept-Stretch-Rationalization model (Bølstad 2020).

7   Conclusion

Aldrich-McKelvey scaling and its adaptations remain under-utilised within political science. They represent some of our best methods for placing both external stimuli such as political parties and leaders and survey respondents on the same scale while corrected for differential item functioning and rationalization bias. With the advent of the blackbox scaling method and the later creation of Bayesian Aldrich-McKelvey scaling, these methods can be combined with anchoring vignettes to achieve comparable position placements over time and space. The potential of these scaling methods does not necessarily need to be constrained to political science. Uniquely among latent variable models and scaling methods, Aldrich-McKelvey scaling retains the same lower dimension in its output, where other scaling methods recover a higher dimensionality. There is little reason to imagine that Aldrich-McKelvey scaling could not be used to scale stimuli in other fields - particularly where respondents can be placed along the same scale.

In this collection of notes, I have sought to improve the understanding of the scaling method. In the first note, I show how QR decomposition can be used to produce more accurate results and retain a larger number of respondents. In the second note, I show that the model is not identified when the number of stimuli is below 3. In the third note, I show how the heteroskedasticity-robustness of the method is not the reason to favour it over taking the mean. Instead, the strong point of the Aldrich-McKelvey method is the fact it makes our assumptions regarding the data generating process transparent and in the longer

term facilitates new adaptations for situations where these assumptions are not met. Finally, I show that the Aldrich-McKelvey scaling method is different in form to the Bayesian Aldrich-McKelvey scaling method and thus in interpretation of the individual parameters.

Bibliography

Aldrich, John H, and Richard D McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71 (1): 111–30.

Armstrong, David, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal. 2020. *Analyzing Spatial Models of Choice and Judgment in r.* 2nd ed. Chapman; Hall/CRC.

Bakker, Ryan, Seth Jolly, Jonathan Polk, and Keith Poole. 2014. "The European Common Space: Extending the Use of Anchoring Vignettes." *The Journal of Politics* 76 (4): 1089–1101.

Bølstad, Jørgen. 2020. "Capturing Rationalization Bias and Differential Item Functioning: A Unified Bayesian Scaling Approach." *Political Analysis* 28 (3): 340–55.

Cikanek, Erin, Kevin McAlister, and Hwayong Shin. 2019. "Ordered Bayesian Aldrich-McKelvey Scaling: Improving Bias Correction on the Liberal-Conservative Scale." *Annual Meeting of the Midwest Political Science Association.*

Goldberg, David. 1991. "What Every Computer Scientist Should Know about Floating-Point Arithmetic." *ACM Computing Surveys (CSUR)* 23 (1): 5–48.

Gouret, Fabian. 2021. "Empirical Foundation of Valence Using Aldrich–McKelvey Scaling." *Review of Economic Design*, 1–50.

Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll, and Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59 (3): 759–74.

Hollibaugh, Gary E, Lawrence S Rothenberg, and Kristin K Rulison. 2013. "Does It Really Hurt to Be Out of Step?" *Political Research Quarterly* 66 (4): 856–67.

Jessee, Stephen A. 2021. "Estimating Individuals' Political Perceptions While Adjusting for Differential Item Functioning." *Political Analysis* 29 (1): 1–18.

King, Gary, Christopher JL Murray, Joshua A Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207.

Lo, James, Sven-Oliver Proksch, and Thomas Gschwend. 2014. "A Common Left-Right Scale for Voters and Parties in Europe." *Political Analysis* 22 (2): 205–23.

Palfrey, Thomas R, and Keith T Poole. 1987. "The Relationship Between Information, Ideology, and Voting Behavior." *American Journal of Political Science*, 511–30.

Poole, Keith T. 1998. "Recovering a Basic Space from a Set of Issue Scales." *American Journal of Political Science*, 954–93.

Poole, Keith T., Jeffrey B Lewis, Howard Rosenthal, James Lo, and Royce Carroll. 2016. "Recovering a Basic Space from Issue Scales in r." *Journal of Statistical Software* 69 (7).

Saiegh, Sebastian M. 2009. "Recovering a Basic Space from Elite Surveys: Evidence from Latin America." *Legislative Studies Quarterly* 34 (1): 117–45.

Satopää, Ville A. 2017. "Combining Information from Multiple Forecasters: Inefficiency of Central Tendency." *arXiv Preprint arXiv:1706.06006.*

Satopää, Ville A, Marat Salikhov, Philip E Tetlock, and Barbara Mellers. 2021. "Bias, Information, Noise: The BIN Model of Forecasting." *Management Science.*