201a æ a a

What's the Deal with Drum Corps International Scoring?

James W. Boudreau[y]
Matthew McKenzie[z]

May 16, 2020

Abstract

The Drum Corps International (DCI) is an organization for drum and bugle corps based out of Indiana. They are liable for developing rules and providing standardized adjudication at DCI sanctioned competitions all throughout the USA and Canada with the goal of crowning a world championship corps at the end of the season. We question whether some categories of scoring are more important than explicitly listed. Speci cally, although there are three main scoring categories that account for 40, 30, and 30 points to add up to a possible 100 points total for each team, the two categories with lower point potentials display higher variations in the magnitude of score differential when judges in those categories do not agree with a team's overall rank. This suggests that if judges do not agree with the rankings of teams by judges in other categories, they may skew the amount of points they award teams within their category to impact the nal ranking.

JEL Classi cation Codes: D70, Z20.
Keywords: Scoring mechanisms, social choice, drum corps.

Preliminary draft and working title. Please do not cite without authors' permission.

[y]Kennesaw State University, Email: jboudre5@kennesaw.edu.
[z]Kennesaw State University, Email: mmcken24@students.kennesaw.edu

1 Introduction

The aggregation of heterogeneous preferences into an ultimate decision is not just important for voting. It's important for the scoring of many of our favorite competitive endeavors. Take the Drum Corps International Competition, for example.

Drum corps international (DCI) is the governing body that sets out rules for DCI competitions. There are two levels of DCI competition corps, open and world class. Open class is the lower level of the two and world class is the highest level of competition. Corps gather in the summer months to develop a routine and starting in June they will tour across the USA and Canada competing against other corps in competition. In August of the same year they gather and compete in the world championship.

In order to determine a winner, the scores for each corps, 8 categories with one to two judges are assigned. The categories are as follows: general effect, music and visual analysis, visual pro ciency, color guard, brass, and ensemble and eld percussion. Scores are provided for each category, but whether or not each component of the score system is as in u(ut)-22o350 (as)r0Druut

Table 1: 2019 Drum Corps International World Championships Scores

	General Effect		Visual		Music		Total	
Team	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Blue Devils	39.3	2	29.775	1	29.25			

Table 2: 2019 Drum Corps International World Championships Scores

General Effect		Visual		Music		Total		
Score	Rank	Score	Rank	Score	Rank	Score	Score Difference	Rank
39.3	2	29.775	1	29.25	3	98.325		1
39.45	1	29.4 5	2	29.338	1	98.238	0.087	2
38.425	4	29.1	3	29.075	4	96.6	1.638	3
38.525	3	28.7	5	1 29.338	1	96.563	0.037	4
38.075	5	28.55	6	28.775	5	95.4	5	1.163
37.5	6	29	4	27.988	6	94.488	6	0.912
36.925	7	27.425	9	27.7	7	92.05	7	2.438
36.225	8	27.65	7	27.35	9	91.225	8	0.825
35.4	10	26.75	10	27.688	8	89.838	9	1.387
35.55	9	27.5	8	26.25	11	89.3	10	0.538
34.85	11	26.575	11	26.125	12	87.55	11	1.75
34.6	12	26.1	12	26.538	10	87.238	12	0.312
								(0.755)

ve and six are each resorted, and we do so in order to sort each scoring category from highest to lowest scores. We then repeat the exercise of looking at score differentials by rank. For example, the difference in the team that ranked rst in the General Effect category as compared to second, second versus third, and so on for each of the three categories. This is presented in Table 3, along with the standard deviation for these differentials presented at the bottom of each of those columns in parentheses.

Though it is a bit strange the standard deviation of score differentials in the General Effect category is a bit smaller than the others given that the General Effect category entails a total possible 40 points while the others have only 30, the three do appear to be fairly similar. This suggests that within any one of the three categories, teams within one rank of each other seem to display similar variation in score.

What shows a bit more discrepancy is if we conduct a similar exercise, but this time look at the score

Table 3: Score Categories Each Sorted by Rank with Score Differentials

General Effect			Visual			Music		
Score	Rank	Differential	Score	Rank	Differential	Score	Rank	Differential
39.45	1							

Table 4: Score Differentials by Category Based on Overall Rank

General Effect			Visual			Music		
Score	Rank	Differential	Score	Rank	Differential	Score	Rank	Differential
39.3	2		29.775	1		29.25	3	
39.45	1	-0.15	29.45	2	0.325	29.338	1	-0.088
38.425	4	1.025	29.1	3	0.35	29.075	4	0.263
38.525	3	-0.1	28.7	5	0.4	29.338	1	-0.263
38.075	5	0.45	28.55	6	0.15	28.775	5	0.563
37.5	6	0.575	29	4	-0.45	27.988	6	0.787
36.925	7	0.575	27.425	9	1.575	27.7	7	0.288
36.225	8	0.7	27.65	7	-0.225	27.35	9	0.35
35.4	10	0.825	26.75	10	0.9	27.688	8	-0.338
35.55	9	-0.15	27.5	8	-0.75	26.25	11	1.438
34.85	11	0.7	26.575	11	0.925	26.125	12	0.125
34.6	12	0.25	26.1	12	0.475	26.538	10	-0.413
		(0.410)			(0.661)			(0.546)

Table 5: Standard Deviations of Score Differentials by Category, Based on Overall Rank, for Four Major DCI Events

Event	General Effect	Visual	Music	Total Score
2019 World Champs	0.410	0.661	0.546	0.755
2019 Semi nals	0.395	0.323	0.320	0.547
2019 Prelims	0.338	0.480	0.468	0.835
2018 World Champs	0.329	0.514	0.527	0.523

3 Estimation and Results

We use a difference-in-difference estimation approach to determine whether or not a team's score in a particular category is signi cantly different on average given that it was ranked differently in that category than in the nal overall ranking of teams for an event. More speci cally, we construct three dummy variables, one for each major scoring category, that take a value of 1 only if a team ranked better in that category than their nal overall ranking in the preliminary, semi- nal, and world championship DCI events of 2019. We then run a treatment effects regression with inverse-probability weighting (IPW), since we are comparing two groups that may not be selected at random. That is, since our data is the result of judgement and we are testing whether the judges gave more disparate scores to teams that performed differently as compared to how other judges ranked them, this is not the same as comparing two groups of randomly selected subjects in a medical trial; clearly there may have been some selection on the part of judges. Fortunately, IPW matching is suited for just this conducting a difference-in-difference comparison in this sort of environment (Angrist and Pischke, 2009; Huber, 2014).

Figures 1-3 present our preliminary results with data from the DCI 2019 World Championship, Semi- nal, and Championship Preliminary events, in addition to the 2018 World Championships.[2] In each regression we are testing whether or not there is a signi cant difference in the average score within a category depending on whether or not a team's rank within that category aligns with their overall (Total) rank or not. We run three separate regressions, one with the teams' General Effect scores as the outcome (dependent) variable, one with their Music scores, and one with visual scores. In each we use the dummy variable for that score category as the treatment variable, and control for the team's ranks in the two other categories.

2

that if a team has a lower rank in the visual category than their overall rank, their actual music score itself is on average 0.305 points lower. In a competition where placements are sometimes decided by margins of much less, this is an interesting finding. Our results are similar for the Visual scoring category, though they are only significant at the 10% level, with a coefficient of -0.23 (see Figure 3). These results are in-line with the analysis of score differentials from the previous section.

4 Discussion

What does it mean that the Visual category seems to score more aggressively when they disagree with other judges? Are the judges in this category attempting to change the outcome of the competition? Obviously we can not possibly say, though a summation-score format does leave such possibilities open regardless of our results here.

Any time judgements must be aggregated or group decisions made, controversy may follow. By further analyzing these types of mechanisms, we hope to open the door to future improvements and alternatives. Gerardi et al. (2009) and Clemens and Puppe (2010) have offered axiomatic bases for judgement aggregation mechanisms. Although the DCI is a nuanced environment, that does not mean it has to be subject to any one category's dictatorship.

The fact that the current scoring system is based on the summation of numerical scores, and that judges have potential leeway to manipulate the magnitude of several teams' scores without necessarily changing how they rank teams within their category while still impacting the overall scores of those teams, is what allows the type of possible manipulation we are suggesting could be present here. In the future we plan to propose an alternative scoring methodology that is based strictly on the rankings of teams within the individual scoring categories. Judges would then still rate teams within each category and rank them, and the scoring method would aggregate the categorical rankings into a final ranking, but the type of manipulation we suggest could be present (but do not say necessarily is present) in the current system would no longer be possible.

Figure 1: Regression Results: Treatment Effects Estimation with Inverse-Probability Weighting for General Effect

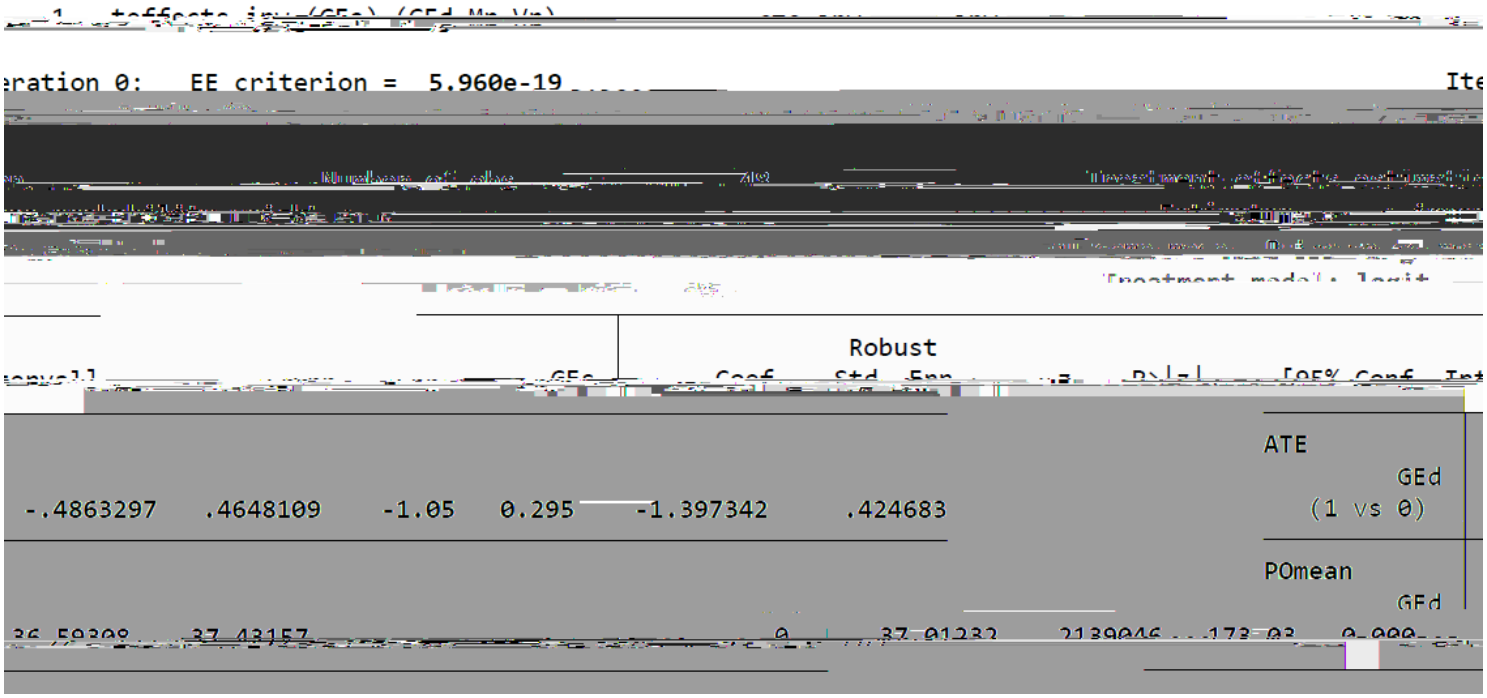

Figure 2: Regression Results: Treatment Effects Estimation with Inverse-Probability Weighting for General Effect