First or Last: An Economic Model of Contest Judging and Early Performance Bias

Tim Barmby[1] (University of Aberdeen) & Ben Kriechel (Maastricht University)

November 2010

Preliminary – Do not quote without authors' consent

Abstract:- It has been observed that in sequential tournaments performing early can reduce an individual's probability of winning even when the order of performance has been randomly assigned. We offer an explanation of this which doesn't depend on extra assumptions about the psychological make-up of judges but depends only on judges' having some uncertainty about their assessment, and that they learn.

I Introduction

Ginsburgh and Van Ours (2003) draw attention to an interesting characteristic regarding the organisation of, in particular, the Queen Elizabeth piano competition. In this competition the order in which entrants perform is randomised, this is done to try and ensure fairness, however it appears there are some unintended consequences of this which we seek to explore in this paper.

The competition then proceeds through three stages, after the first stage the number of competitors is reduced to 24, after the second stage to 12. The third stage selects the winner. Ginsburgh and Van Ours make the observation *"In the last two stages, members of the jury grade candidates after every day of performance. Marks are*

[1] Corresponding author tim.barmby@abdn.ac.uk

given without discussion between the judges and cannot be changed after having been turned in"

This means that in the final stage, where 12 finalists perform over 6 evenings, judges are submitting scores for the first two performers, before they hear the last 10 (which they will over the next 5 evenings). Ginsburgh and Van Ours, find as an empirical matter the order of performance has a significant negative effect on the probability of success, even though the order has been selected randomly and so cannot be related to ability.

This aspect of the effect of ordering in sequential tournaments has been recognised elsewhere, Bruine de Bruin (2006) looks at the same question in the context of ice-skating competitions, her explanation is based in the psychology of the judges. Judges place more weight on new aspects of performance, therefore it is more difficult for early performers to gain by this sort of comparison, Page and Page (2010) look at the same effects in television talent contests.

In this paper we consider a possible theoretical basis for this bias, which is based on how judges might deal with their own uncertainty. The next section starts to build a model of way the judges may behave.

II Judging

There are two types of agents, judges and competitors. Competitors, are defined by "ability" μ, in the model we construct this can be high, H, medium, M, or low, L. Judges are the functional relationship between μ, ability, and the score q attached to the competitor's performance. We wish to capture in our model the idea that judges are better able to identify high and low ability, the problem comes with the medium ability. So one joint probability distribution of ability and score which characterises a judges behaviour, could be written as

Table 1: Joint distribution of ability and score

		Score, Q			
Ability, μ		H	M	L	
	H	a	0	0	a
	M	b/3	b/3	b/3	b
	L	0	0	1-a-b	1-a-b
		a+b/3	b/3	1-a-2b/3	1

This table is equivalent to the output function in a standard tournament, as it serves to describe the probabilistic relationship between output and ability. This judge described in this table might be considered to be not a very good judge, for while he/she knows what is good, but the variability of reporting in the mid range is high. This is an extreme example for the purposes of exposition, for an interesting study on the reliability of judging see Hodgson (2008). One thing which we will assume is that judges get better, we will return to this idea later.

We can see in the above that the judges uncertainty regarding the medium ability performers results in too higher proportion being reported as high ability, a+b/3 instead of a. The nature of his uncertainty of course means that he can't tell which in H are truly H and which are M that he has mistakenly thought are H, or indeed which are L and which are M that he has mistakenly thought L. To counter this he may decide to revise downward a proportion of the H's and upwards a proportion of the L's. Let us consider that he revises a proportion c, of both the H's and the L's and reports them as M.

Table 2: Joint distribution of ability and reported score

		Reported Score, R			
Ability, μ		H	M	L	
	H	a(1-c)	ac	0	a
	M	b(1-c)/3	b/3+2bc/3	b(1-c)/3	b
	L	0	(1-a-b)c	(1-a-b)(1-c)	1-a-b
		(a+b/3)(1-c)	c+b/3-bc/3	(1-a-2b/3)(1-c)	1

Here we are making the distinction between score *Q*, what the judge has in his head, and what he chooses to report, *R*. Intuitively if there is a high proportion of M's then his revisions are more likely to be correct. This can be established by writing the matrices

$$Q = \begin{bmatrix} a & 0 & 0 \\ b/3 & b/3 & b/3 \\ 0 & 0 & 1-a-b \end{bmatrix} \qquad R = \begin{bmatrix} a(1-c) & ac & 0 \\ b(1-c)/3 & b/3+2bc/3 & b(1-c)/3 \\ 0 & (1-a-b)c & (1-a-b)(1-c) \end{bmatrix}$$

It will also be useful to think of the transformation which takes *Q* to *R* we define the matrix *C* which describes the judge being cautious in his assessment by 'biasing' his reports towards M.

$$C = \begin{bmatrix} 1-c & c & 0 \\ 0 & 1 & 0 \\ 0 & c & 1-c \end{bmatrix}$$

And note that $R = Q.C$.

The matrix trace of *Q* and *R* give the probability of making the right decision, firstly when the judge reports his true assessment, and then when he/she is cautious. It is straightforward to show that $\mathrm{tr}(Q) < \mathrm{tr}(R)$ when $b>3/5$. This result suggests that it when there is a "lot" of uncertainty it will a good idea to be cautious.

III Sequential performances

The situation of most interest to us is where contestants perform in sequence as this corresponds to the situation of piano competitions we took as the start of our paper, and judges give a report on the performance immediately after each performance.

Here, we will argue, judges will be concerned with two things, assessing performers absolute performance, and also their relative performance. In the simplest two contestant situation the possible outcomes are

Table 3: Possible configuration in a two round contest.

		Score								
		H,H	H,M	H,L	M,H	M,M	M,L	L,H	L.M	L,L
Ability	H,H	(=,=)	(=,+)	(=,+)	(=,-)	(=,=)	(=,+)	(=,-)	(=,-)	(=,=)
	H,M	(+,=)	(+,+)	(+,+)	(+,-)	(+,=)	(+,+)	(+,-)	(+,-)	(+,=)
	H,L	(+,=)	(+,+)	(+,+)	(+,-)	(+,=)	(+,+)	(+,-)	(+,-)	(+,=)
	M,H	(-,=)	(-,+)	(-,+)	(-,-)	(-,=)	(-,+)	(-,-)	(-,-)	(-,=)
	M,M	(=,=)	(=,+)	(=,+)	(=,-)	(=,=)	(=,+)	(=,-)	(=,-)	(=,=)
	M,L	(+.=)	(+,+)	(+,+)	(+,-)	(+.=)	(+,+)	(+,-)	(+,-)	(+.=)
	L,H	(-,=)	(-,+)	(-,+)	(-,-)	(-,=)	(-,+)	(-,-)	(-,-)	(-,=)
	L,M	(-,=)	(-,+)	(-,+)	(-,-)	(-,=)	(-,+)	(-,-)	(-,-)	(-,=)
	L,L	(=,=)	(=,+)	(=,+)	(=,-)	(=,=)	(=,+)	(=,-)	(=,-)	(=,=)

If the judge chooses simply to report the score in his head each round then the probability of the outcomes in this table can be computed as $Q \otimes Q$

$$Q \otimes Q = \begin{bmatrix} a^2, & 0, & 0, & 0, & 0, & 0, & 0, & 0, & 0 \\ \frac{a\,b}{3}, & \frac{a\,b}{3}, & \frac{a\,b}{3}, & 0, & 0, & 0, & 0, & 0, & 0 \\ 0, & 0, & a(1-a-b), & 0, & 0, & 0, & 0, & 0, & 0 \\ \frac{a\,b}{3}, & 0, & 0, & \frac{a\,b}{3}, & 0, & 0, & \frac{a\,b}{3}, & 0, & 0 \\ \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9}, & \frac{b^2}{9} \\ 0, & 0, & \frac{b(1-a-b)}{3}, & 0, & 0, & \frac{b(1-a-b)}{3}, & 0, & 0, & \frac{b(1-a-b)}{3} \\ 0, & 0, & 0, & 0, & 0, & 0, & a(1-a-b), & 0, & 0 \\ 0, & 0, & 0, & 0, & 0, & 0, & \frac{b(1-a-b)}{3}, & \frac{b(1-a-b)}{3}, & \frac{b(1-a-b)}{3} \\ 0, & 0, & 0, & 0, & 0, & 0, & 0, & 0, & (1-a-b)^2 \end{bmatrix}$$

The first question to ask is whether it is better to be cautious in this two round contest and also when to be cautious. We can show that tr($QC \otimes Q$) – tr($Q \otimes Q$) = $\frac{c(5b-3)(3-2b)}{9}$ which will be positive when b >3/5, as before, but tr($Q \otimes QC$) – tr($Q \otimes Q$) will also equal the same, it doesn't seem to matter in which round you are cautious (indeed you can improve your probability of making the right decision by being cautious in both rounds)[2].

What is missing? We argue in the next section that the missing ingredient is (even a small amount of) learning by judges.

IV Learning

We noted in section II that a realistic aspect of this situation that we might wish to capture is that judges improve. So whatever the nature of the contest judges will get better, even very slightly, each performer they observe. So in the context of piano contests each piece listened to improves a judge's ability. We can model this by assuming a learning matrix *L*

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -d & 1+2d & -d \\ 0 & 0 & 1 \end{bmatrix} \qquad QL = \begin{bmatrix} a & 0 & 0 \\ b(1-d)/3 & b(1+2d)/3 & b(1-d)/3 \\ 0 & 0 & (1-a-b) \end{bmatrix}$$

d can be considered arbitrarily small, capturing a slow accumulation of experience. If Q represents the judge's ability to assess performers in round 1, QL is there ability at round 2.

Without learning it won't matter whether the judge biased his assessment down in the first round or the second, but with learning it does.

[2] All computations are performed in Maple

$$tr(QC \otimes QL) - tr(Q \otimes QL) = \frac{c(5b-3)(2bd-2b+3)}{9}$$

Which will be positive if *b* > *3/5*, and *d* > 0 that is there is some learning, but it can also be shown that

$$tr(QC \otimes QL) - tr(Q \otimes QLC) = \frac{2b^2cd}{3}$$

That is it is better to be cautious in the first round, this give the basis for the early performance bias observed in piano and other sequential tournaments.

V Conclusion

This short note describes a theoretical reason for early performance bias which does not depend on extra assumptions about the psychological make up of judges but depends only on judges' having some uncertainty about their assessment, and additionally that judges learn, even very slowly.

References

Bruine de Briun W (2006) "Save the last dance II: Unwanted serial position effects in figure skating judgements" *Acta Psychologia* 123, pp 299-311
Ginsburgh V A and J Van Ours (2003) "Expert Opinion and Compensation: Evidence from a Musical Competition" *American Economic Review* Vol 93:1 pp 289-295
Hodgson R T (2008) "An Examination of Judge Reliability at a Major US Wine Competition" *Journal of Wine Economics* Vol 3:(2) pp 10
Page L and K Page (2010) "Last shall be first: A field study of biases in sequential performance evaluation on the Idol series" *Journal of Economic Behaviour and Organisation* 73 pp 186-98.