1 Introduction

Contest success functions (CSFs) are an essential part of the economic analysis of contests, as they describe the relationship between the efforts participants invest in a contest and their consequent chances of winning. In the literature two major types of CSF have emerged. The first (and most popular) type relates the probability of success to the relative efforts participants exert in a contest. This type is commonly referred to as the Tullock CSF, after Tullock (1980). In the second class of CSFs the absolute difference of participant efforts determines their probability of success. This class is commonly referred to as difference–form or Hirshleifer CSFs, because Hirshleifer (1989) introduced them. Several theoretical contributions discuss the use of both types of CSFs and their application in more detail (see e.g. Skaperdas (1996), Hirshleifer (1989) and Alcalde & Dahm (2007)). However, the empirical question whether real-life contests behave according to either one of these CSFs remains unanswered.

To the best of my knowledge this note presents the first empirical assessment of Tullock and Hirshleifer CSFs, using real-life contests. Sports leagues offer an excellent testing ground for contest theory, as they provide a large number of contests between different parties with a fixed set of rules. In this sense the "technology" of the contest and with it the CSF remains unchanged. I rely on a dataset containing over 65.000 fixtures from the American major sports leagues (NFL, MLB, NBA and NHL) to estimate both classes of CSF and compare their fit. I find that Tullock CSFs fit the data better than Hirshleifer CSFs in all sports and for all tested models. The fit of the Tullock models is significantly better for the NBA, NFL and in one case for the MLB, while the difference in the NHL falls short of being significant. A second observation from the estimates is that home-advantage results in asymmetric CSFs in all sports. This means home teams have to put in less effort (measured as player wage expenditures) to obtain a similar probability of success than their visiting counterparts. These results in general confirm the CSFs used in most economic models of sports (see Szymanski (2003)).

A first section of this note contains some background on the Tullock and Hirshleifer CSFs I estimate. Then section 3 describes the dataset. Section 4 finally presents the empirical results.

2 Theory

2.1 Tullock's ratio-form CSF

The ratio-form CSF I estimate goes back to the seminal paper by Tullock (1980). Skaperdas (1996), Clark & Riis (1998) and Kooreman & Schoonbeek (1997) have provided a sound theoretical foundation for this functional form, using different axioms to derive it. It has also been the most popular CSF in the economic analysis of sports (see Szymanski (2003) for more on this). For my application I specify the probability of the home team winning as

$$\Pr(win = 1\,|I_h, I_a) = \frac{\alpha I_h^{\beta_h}}{\alpha I_h^{\beta_h} + I_a^{\beta_a}} \tag{1}$$

I_h and I_a represent the efforts (investments) of the home and away team respectively. Both β parameters indicate the return on investment in terms of winning probability, with high β's indicating higher returns. The α is a measure for the asymmetry of the contest. A larger α means home-advantage plays a more important role. I estimate four different forms of (1). In model (a) I restrict the contest to be symmetric, meaning that $\alpha = 1$ and $\beta_h = \beta_a$. Under (b) I allow for asymmetry by giving up the equality of both β's, keeping $\alpha = 1$. In this case I expect to find $\beta_a < \beta_h$, if home advantage is present. Under model (c) I allow for $\alpha \neq 1$, but keep

$\beta_h = \beta_a$. In this case home advantage would mean to find $\alpha > 1$. Finally, model (d) presents the unrestricted estimation of (1).

2.2 Hirshleifer's difference-form CSF

In order to parameterize a difference-form CSF it is necessary to scale an absolute difference in efforts (which may have any size) into a probability with boundaries 0 and 1. I follow the seminal paper from Hirshleifer (1989) by choosing the logit transformation. Furthermore, Skaperdas (1996) shows that the logit transformation difference-form CSF is the only one which may be theoretically founded in a similar way as the Tullock CSF. For my model the Hirshleifer CSF is given by

$$\Pr(win = 1 \,| I_h, I_a) = \frac{1}{1 + \exp(-\alpha - \beta_h I_h + \beta_a I_a)} \tag{2}$$

where I_h and I_a are again the investments of the home and away team. The parameters α, β_h and β_a may be interpreted in a similar fashion as before. As in the Tullock case I estimate four distinct forms, in (a) $\alpha = 0$ and $\beta_h = \beta_a$, in (b) $\alpha = 0$, but $\beta_h \neq \beta_a$, under (c) $\alpha \neq 0$ and $\beta_h = \beta_a$ and finally (d) is the unrestricted model. Here the existence of home-advantage means α should be larger than 0, or $\beta_h > \beta_a$.

3 Data

In order to estimate the parameters of the models (a)-(d) for both CSFs I construct a dataset on the 4 American major leagues (NFL, MLB, NBA and NHL). I consider the investments teams made in playing talent, measured by their total payroll, as the effort invested in winning matches. Data on payrolls for all major leagues come from the online database of the newspaper USAToday[1]. I convert payroll data to 2009 US dollars using monthly CPI statistics from the US department of Labor. Data on sports results were taken from the online archive shrpsports.com[2]. Getting home advantage in the playoff games depends on previous results, which makes it endogenous. Therefore I restrict attention to regular season fixtures. I also abstract from fixtures ending in a tie, because these do not produce clear-cut outcomes. This means I only include NHL results after 2005, because up until that point matches were allowed to (and often did) end in a tie. For the other sports I drop tied matches, but these constitute a minor amount of data.[3] Table 1 presents summary statistics of the dataset. The variable *home_res* refers to the home team's result measured as 0 for a loss and 1 for a win, while *home_pay* and *away_pay* are the payroll data. Notice that the average home result is above 0.5, which is already suggestive of the fact that home advantage is present in all leagues.

[1] see for example http://content.usatoday.com/sports/basketball/nba/salaries/default.aspx for the NBA data.
[2] online retrievable e.g. for the NBA via http://www.shrpsports.com/nba/teamseas.htm
[3] 2 NFL, 0 NBA and 0 MLB observations were dropped.

league	variable	mean	std dev	min	max	obs	season
NFL	home_res	0.5700	0.4952	0	1	2542	2000-2009
	home_pay	9.38e+07	1.86e+07	5.20e+07	1.54e+08	2542	
	away_pay	9.38e+07	1.86e+07	5.20e+07	1.54e+08	2542	
MLB	home_res	0.5393	0.4985	0	1	50074	1988-2009
	home_pay	6.14e+07	3.34e+07	1.07e+07	2.28e+08	50074	
	away_pay	6.14e+07	3.34e+07	1.07e+07	2.28e+08	50074	
NBA	home_res	0.6051	0.4888	0	1	9717	1999-2009
	home_pay	6.44e+07	1.38e+07	2.56e+07	1.22e+08	9717	
	away_pay	6.44e+07	1.38e+07	2.56e+07	1.22e+08	9717	
NHL	home_res	0.5563	0.4969	0	1	6150	2005-2010
	home_pay	4.60e+07	9094032	2.06e+07	6.74e+07	6150	
	away_pay	4.60e+07	9094032	2.06e+07	6.74e+07	6150	

Table 1: summary statistics

4 Empirical Results

A straightforward way to estimate the parameters of (1) and (2) is to use maximum likelihood estimation. A first step is then to obtain the likelihood functions implied by both CSFs. Assuming match results are independent draws, the likelihood for the Tullock case is:

$$\mathcal{L}_{tul} = \prod_{i=1}^{n} \left(\frac{\alpha I_{hi}^{\beta_h}}{\alpha I_{hi}^{\beta_h} + I_{ai}^{\beta_a}} \right)^{y_i} \left(\frac{I_{ai}^{\beta_a}}{\alpha I_{hi}^{\beta_h} + I_{ai}^{\beta_a}} \right)^{1-y_i} \quad (3)$$

where y_i is the dependent variable *home_ res*. Likewise the Hirshleifer CSF leads to the standard logit likelihood:

$$\mathcal{L}_{hir} = \prod_{i=1}^{n} \left(\frac{1}{1+\exp(-\alpha - \beta_h I_{hi} + \beta_a I_{ai})} \right)^{y_i} \left(\frac{\exp(-\alpha - \beta_h I_{hi} + \beta_a I_{ai})}{1+\exp(-\alpha - \beta_h I_{hi} + \beta_a I_{ai})} \right)^{1-y_i} \quad (4)$$

After taking logarithms of (3) and (4), maximization is carried out using standard econometric software.

Table 2 gives an overview of the estimation results for the Tullock and Hirshleifer models (a)-(d) with computed standard errors in parentheses. Clearly, all coefficients are estimated with their expected sign and size, i.e. a positive β_h, β_a and α in the Tullock model and a positive β_h and α and negative β_a for Hirshleifer. The fact that all α's in model (c) are significantly larger than 1 (or 0 for Hirshleifer) is a strong indicator of the asymmetric nature of sports, favoring home teams. However, the estimates of model (b) are not significantly different in all cases. Another interesting point to see is that success in the MLB is less sensitive to investment than success in the other leagues, as its lower estimated β's suggest. The NFL and NHL offer the largest surplus winning per invested dollar. To get a full picture however one would need to compute marginal effects at each point. Home-advantage appears to be most important in the NBA and least in the MLB. Finally, the reported likelihood values suggest that moving from models (b) and (c) to (d) in most cases adds little explanatory power.

League		Tullock				Hirshleifer (logit)			
		a	b	c	d	a	b	c	d
NFL	$\beta_{h(=a)}$	1.0526	1.0806	1.0729	1.0631	1.05e-08	1.20e-08	1.07e-08	1.07e-08
obs:		(0.207)	(0.209)	(0.209)	(0.242)	(2.19e-09)	(2.20e-09)	(2.21e-09)	(2.57e-09)
2542	β_a		1.0650		1.0827		-9.08e-09		-1.10e-08
			(0.209)		(0.243)		(2.20e-09)		(2.56e-09)
	α			1.3318	1.9091			0.2836	0.3141
				(0.054)	(8.649)			(0.040)	(0.2365)
likelihood		-1749.78	-1723.24	-1723.23	-1723.22	-1750.36	-1725.73	-1724.84	-1724.83
MLB	$\beta_{h(=a)}$	0.2809	0.2868	0.2824	0.2845	4.28e-09	5.35e-09	4.36e-09	4.42e-09
obs:		(0.016)	(0.016)	(0.016)	(0.019)	(2.44e-10)	(2.53e-10)	(2.50e-10)	(2.99e-10)
50074	β_a		0.2779		0.2802		-3.21e-09		-4.31e-09
			(0.016)		(0.019)		(2.53e-10)		(2.96e-10)
	α			1.1717	1.0856			0.1585	0.1520
				(0.011)	(0.356)			(0.009)	(0.022)
likelihood		-34553.2	-34397.5	-34397.5	-34397.5	-34553.8	-34422.5	-34397.9	-34397.9
NBA	$\beta_{h(=a)}$	0.7515	0.7986	0.7867	0.7131	1.03e-08	1.34e-08	1.07e-08	9.87e-09
obs:		(0.070)	(0.072)	(0.072)	(0.092)	(1.06e-09)	(1.07e-09)	(1.08e-09)	(1.54e-09)
9717	β_a		0.7745		0.8614		-7.10e-09		-1.18e-08
			(0.072)		(0.092)		(1.07e-09)		(1.51e-09)
	α			1.5410	22.153			0.4269	0.5584
				(0.032)	(45.47)			(0.021)	(0.137)
likelihood		-6677.04	-6458.16	-6457.95	-6457.35	-6687.84	-6477.45	-6469.28	-6468.81
NHL	$\beta_{h(=a)}$	0.9938	1.0129	1.0066	0.9849	2.24e-08	2.48e-08	2.27e-08	2.28e-08
obs:		(0.109)	(0.110)	(0.110)	(0.130)	(2.48e-09)	(2.49e-09)	(2.49e-09)	(3.07e-09)
6150	β_a		0.9999		1.0291		-2.00e-08		-2.33e-08
			(0.110)		(0.132)		(2.49e-09)		(3.08e-09)
	α			1.2576	2.7372			0.2260	0.2537
				(0.033)	(6.847)			(0.026)	(0.161)
likelihood		-4218.96	-4179.43	-4179.37	-4179.21	-4220.70	-4182.66	-4181.14	-4181.07

Table 2: estimation results for model Tullock (a-d) and Hirshleifer (a-d)

The common likelihood ratio test is inappropriate to assess the fit of most of the models I estimate, because they are non-nested with the naive model containing only a constant term. Therefore I apply the alternative test from Vuong (1989) for non-nested models. In the first columns of table 3 I compare the Tullock models to a naive model, where the average home result is used as a predictor. A positive value of the Vuong statistic implies here that the Tullock model performs better, the p-values indicate significance. Models (b) through (d) significantly outperform the naive model in all cases, whereas model (a) in most cases performs worse. The last columns of table 3 contain the Vuong results when comparing the Tullock to the Hirshleifer models. These results clearly favour the Tullock models over all. They are most significant for the NBA, followed by the NFL and MLB. For the NHL no result reaches the 0.1 level of significance.

League	Model	Tullock vs. constant Vuong	p-value	Tullock vs. Hirshleifer Vuong	p-value
NFL	a	-1.3781	0.1544	1.7180	0.0911
	b	2.6025	0.0135	1.6916	0.0954
	c	2.6050	0.0134	1.7492	0.0864
	d	2.6061	0.0134	1.7653	0.0840
MLB	a	0.0177	0.3989	0.0741	0.3978
	b	8.8594	0.0000	2.4743	0.0186
	c	8.8624	0.0000	0.0614	0.3982
	d	8.8634	0.0000	0.0580	0.3983
NBA	a	-6.7378	0.0000	4.7570	0.000
	b	5.5091	0.0000	5.0000	0.000
	c	5.5313	0.0000	4.8918	0.000
	d	5.5727	0.0000	5.0801	0.000
NHL	a	0.3781	0.3714	1.0728	0.2244
	b	4.7493	0.0000	1.5453	0.1209
	c	4.7554	0.0000	1.0834	0.2218
	d	4.7649	0.0000	1.1420	0.2078

Table 3: Vuong test results Tullock vs. constant and Hirshleifer models

Figure 1 graphically depicts the estimation results of models (a) and (c) for the NBA[4], where the investment of the away team is fixed to be the sample average. Both Tullock curves predict zero win probability for zero investments, whereas the Hirshleifer curves result in a predicted probability around 0.4 at this point. Consequently, the Tullock curves rise more sharply in the lower segment. Around the point of equal strength both curves approach each other, but in the higher segment the Tullock curves increase win probability at a lower pace. Comparing between the curves of model (a) and (c) it is clear that allowing for asymmetric CSFs pushes up both curves along the entire interval. Again, this is indicative of the presence of home-advantage.

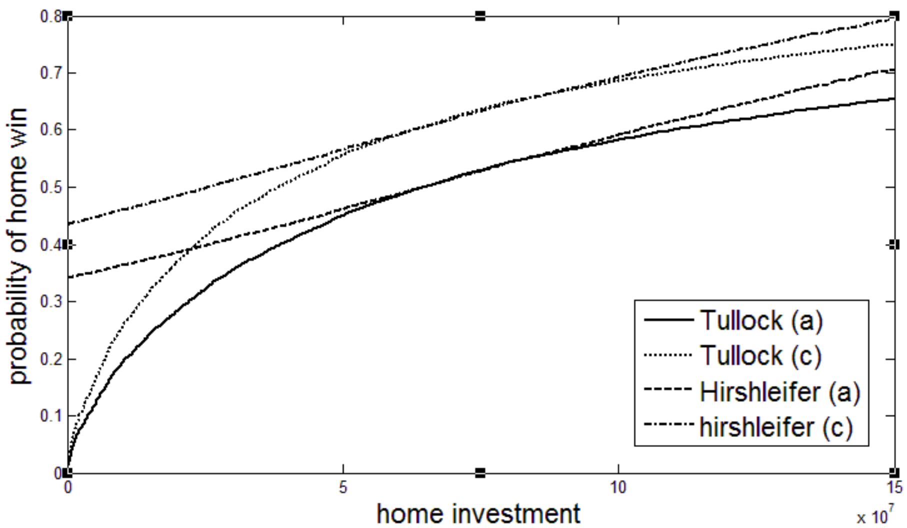

Figure 1: Tullock and Hirshleifer CSFs (a) and (c) for NBA data

[4]Similar figures could have bee created for the other leagues and models.

5 Acknowledgements

I would like to thank Stefan Késenne, Stefan Szymanski and Jan Bouckaert for useful comments and suggestions.