

Norbert Bach, Oliver Gürtler, Joachim Prinz*

Incentive Effects in Tournaments with Heterogeneous Competitors – an Analysis of the Olympic Rowing Regatta in Sydney 2000**

A large part of the theoretical tournament literature argues that rank-order tournaments only unfold their incentive effects if the contestants all have similar prospects of winning. In heterogeneous fields, the outcome of the tournament is relatively clear and the contestants reduce their effort. However, empirical evidence for this so-called contamination hypothesis is sparse. An analysis of 442 showings at the Olympic Rowing Regatta in Sydney 2000 gives evidence that oarsmen spare effort in heterogeneous heats. This implies that competition among staffs with heterogeneous skill levels does not bring about the intended effort levels. However, a separate subgroup analysis shows that only the tournament favourites hold back effort whereas underdogs bring out their best when competing against dominant rivals. A heterogeneous tournament could then be enriched by absolute performance standards to increase incentives of the favourites.

Key words: tournaments, heterogeneity, incentive effects, effort

* Univ.-Prof. Dr. Norbert Bach, Department of Management and Organization, Ilmenau University of Technology, Helmholtzplatz 3, D – 98684, Ilmenau, Germany.
E-mail: norbert.bach@tu-ilmenau.de.

PD Dr. Oliver Gürtler, Department of Economics, BWL II, University of Bonn, Adenauerallee 24-42, D – 53113 Bonn, Germany. E-mail: oliver.guertler@uni-bonn.de.

PD Dr. Joachim Prinz, Department of Management, University of Paderborn, Warburgerstr. 100, D – 33098 Paderborn, Germany. E-mail: joachim.prinz@wiwi.upb.de

** We thank two anonymous reviewers for helpful comments to improve the manuscript and Dipl.-Kfm. Stefan Krämer for research support.

Article received: December 12, 2008

Revised version accepted after double blind review: March 30, 2009.

1. Introduction

As has been shown by Lazear/Rosen (1981), rank-order tournaments work, under certain conditions, as optimal labour contracts yielding the first best level of effort in an environment characterised by moral hazard. Hence, tournaments are not only of academic interest, but quite prominent in day-to-day work environments. For example, co-workers compete for promotions in internal labour markets (Malcomson 1984; Rosen 1986; Lazear 1989; Baker et al. 1994), teams compete for a project, firms compete for contracts. Furthermore, firms set bonus pools that are divided among staffs according to their relative performance (Kräkel 2002; Rajan/Reichelstein 2006), e.g. in sales forces. Formally, such competition for bonuses also can be modelled as a tournament.

Tournament literature suggests that competition will only have incentive effects if an increase in effort increases the chances of winning. In heterogeneous fields, the underdog quickly realises his minimal chances of winning and reduces his effort to the level that just about secures the achievable rank. Vice versa, the favourite anticipates a noticeable competitive edge early on in the tournament. This implies two strategies. First, he also holds back effort right from the start. Second, he reduces his effort as soon as he realises the slowdown of his inferior competitors. The favourite will only increase his effort if he can expect to thereby increase his chances of winning. In this line of argument, heterogeneous tournaments do not have incentive effects; other incentive systems are more likely to affect employees (Frick et al. 2008: 385ff.). As a general conclusion, the incentive to put in additional effort decreases with increasing heterogeneity among competitors.

Despite its prominent role in the theoretical tournament literature, non-experimental empirical evidence for the so-called contamination hypothesis is only limited. In their analysis of PGA-golf tournaments Ehrenberg/Bognanno (1990) did not find unambiguous negative impact of competitor heterogeneity on incentive effects. Independent from their own score, a match partner with a higher player score (exempt players) leads to an underperformance of golf players. For below-average players (nonexempt players), this is in line with the literature. On the other hand, above-average players should be challenged and motivated by the additional competition. However, this could not be confirmed in the study by Ehrenberg/Bognanno (1990). Horse race studies as e.g. provided by Lynch (2005) confirm that a closer competition motivates contestants to higher effort levels. Riders put in more effort in the second half of a race if the margin to the next best rank is only small; i.e. the rider hopes to win a rank by increasing his effort. Finally, with the help of data from the German soccer league, Frick et al. (2008) show that in narrow matches players are more often sanctioned with yellow and yellow/red cards. This indicates that equal matches are played with a higher intensity.

This paper extends the empirical literature on the contamination hypothesis by analysing competitor heterogeneity, effort levels and results of the Olympic Rowing Regatta in Sydney 2000. We analyse results (times, splits, and ranks) published by the world rowing association FISA in combination with information regarding the competing teams (age, experience, rank in the 1999 world championship). As a heteroge-

neity measure we use the ordinal variable tournament stage, i.e. heat, repechage, semi-final, final. The analysis shows that with progression in the tournament, i.e. decreasing competitor heterogeneity, the oarsmen row significantly faster times. This confirms undoubtedly that heterogeneous line-ups have smaller incentive effects than close competition. Therefore, principals in internal labour-markets should strive for homogeneous competitor fields when setting up internal rank-order tournaments.

Furthermore, we present the first field data analysis of differences between efforts shown by favourites and by underdogs. So far, this has only been studied in experiments with students (e.g. Schotter/Weigelt 1992). The analysis of the single sculling events in Sydney 2000 shows that only favourites hold back effort whereas underdogs predominantly row sports-physiologically optimal race strategies. As a result, firms organising heterogeneous tournaments have to find ways to reinstall incentives for the favourites. One alternative would be to handicap favourites to make competition more even (e.g. Meyer 1991). Handicaps, however, entail serious problems. They may, for instance, be at odds with regulations from labour law forbidding worker discrimination. Presumably, a better way to keep favourites' incentives high is to enrich the tournament with absolute performance standards (Clark/Riis 2001). To be more concrete, the size of the winner prize may depend on the winner's absolute performance (i.e. on whether the winner's performance is above some standard). Then, favourites have an incentive to put forth effort even if they are far ahead of their competitors since slacking off may come at the risk of not meeting the performance standard.

In our interpretation, for underdogs already the participation in a tournament of prime importance – here the Olympic Games – has a very high incentive effect. This implies that in internal labour markets the organiser has to point out the relevance of the tournament; not surprisingly, management attention is expected to be a key motivational factor. Furthermore, considering the specific incentive of participating in Olympic Games, rank-order tournaments will only unfold their positive incentive effect if they are not carried out too often. If a homogeneous competitor field can not be achieved, other incentive schemes are more likely to affect employees. In a heterogeneous internal labour market, rank-order tournaments should only be held if a minimum incentive through selection for participation in the tournament is guaranteed.

The remainder of the paper is organised as follows. In the next section we introduce our empirical setting, the operationalisation of the contamination hypothesis, and the available data. Section 3 presents the estimation models and empirical results. In section 4 we analyse separately the subgroups of favourites and underdogs. The paper ends with a discussion of the results and an outlook on future work.

2. Hypotheses and empirical setting

The following analysis focuses on the effort competing rowing teams show depending on the heterogeneity of the field. As a heterogeneity measure we use the achieved tournament stage. Because of the regional qualification modus of the Olympic Games, the fitness and skill levels among the contenders vary between multiple world champions and starters who would not qualify for a national final in a strong rowing nation

like e.g. Great Britain. Similar to other sporting contests, in the first round (heats), medal contenders compete against underdogs in heterogeneous fields. In each following tournament stage the line-ups are selected by the results in the preceding stage. The aim of this regulation is to form homogeneous line-ups for the final tournament stage (Olympic Rowing is a full rank tournament with finals A, B, C, and D); the final A consisting of the best six teams.

The effort of the rowing teams is measured by the end time to finish the Olympic 2.000m distance. Rowing times are strongly affected by weather conditions, and to a lesser extent by water temperature and water depths. Therefore, the FISA does not recognise world records, but only world best times rowed on courses that fulfil the FISA requirements. Most of these best times have been achieved with a strong tailwind, warm water temperature, and in deep water. Therefore, rowing experts only discuss absolute times in the context of the local conditions. However, for the Olympic Rowing Regatta in Sydney 2000 the weather conditions have been documented by the Australian Institute for Sports as favourable and stable over all days of the competition (Kleshnev 2001). This allows to specify the contamination hypothesis for Olympic Rowing as follows:

Hypothesis 1: Rowing teams row faster times with every progression in the tournament.

For a team that has qualified for the final this hypothesis indicates that the time in the final is faster than the time in the semi-final; the time in the semi-final is faster than the time in the heat.

The data used in this study has been compiled from the results and athletes databases hosted by the FISA (www.fisa.org). In order to avoid a distortion by inferior contenders from non-rowing nations, we focus on crews that finished in the top 12 ranks, having rowed in the final A or final B. The information analysed here comprises bibliographic data on 317 male and 183 female athletes from 44 nations. Race information covers results of 173 teams (103 male, 70 female) competing in 14 different events, rowing in 6 different boat types (single sculls, double sculls, quadruple sculls, pair, four, and eight). The performance was measured by the finishing times to complete the 2.000m rowing course, and split times for each four 500m quarters.

Because of the specific tournament structure (full rank-order tournament), for each team data is available for different tournament stages (heat, repechage or semi-final, final). Hence, the data could be ordered in form of a “balanced panel”, where the unit of analysis is the progression level of the different teams. Note in this respect that in some events heat winners are directly qualified for the final, i.e. not all teams had to row semi-finals. All in all, this results in a total number of cases of $N = 442$ (173 teams, each rowing 2 or 3 tournament stages). For each race we know the respective end time, split times, and tournament stage. This allows a straight test of the above derived hypothesis: Teams row faster finishing times with their advancement in the tournament. Furthermore, the panel character of the data allows the use of estimation methods accounting for unknown team specific variables (e.g. boat quality, team coordination, physical fitness, etc.) that may affect the dependent variable (Kahn 1993).

In addition to the tournament stage (HET) we control for further covariates that may explain the variance of our endogenous parameter (finishing times). Known to be of importance in endurance sports are variables that describe the physical strengths of the athletes. As a first approximation, we use a team's average age (AGE), and its average race experience (EXP) as indicators. Race experience is measured by the number of years between their first participation in a world championship, a world-cup regatta, or Olympic Games, and the Sydney 2000 regatta. Therefore, an athlete who never competed at one of these international regattas before the Sydney Games is coded as inexperienced (=0). The positive effect of experience may be reduced by an aging component (Fair 1994; Maxcy 1997; Hübl/Swieter 2002). Therefore, we additionally include a squared experience term (EXP_2) in the estimation model. Probably the best estimate of team quality is the rank achieved in the previous 1999 world championship event (WM99); this term is also included in the estimation model. Since a better rank at the 1999 world championship indicates a stronger team, we expect some "path dependence" and hence a positive effect on the finishing times.

One drawback of our database might be, that it is comprised of pooled information on boat categories and athlete sex. Both variables are however expected to have an effect on the dependent variable. For example, given the coordination necessary in crew boats, we expect quicker and easier observable changes in racing strategy in the single sculls events. Similarly, because of comparable physiological capacity of the athletes, we expect smaller differences in speed in the lightweight categories. Therefore, we include categorical variables (SEX, BOAT, LW) in the estimation model to control for these important effects.

Last but not least, end time is primarily determined by the number of oarsmen in the boat; eights are faster than singles. Originating from calculations in the former German Democratic Republic, the sport of rowing has a long tradition to account for these speed differences in comparing relative times across events. The absolute end time is set in relation to a reference time, the so-called "gold standard". Rowing coaches calculate these "gold standards" as extrapolations of preceding world championships and world-cup regattas (Kleshnev 2001; Teti/Nolte 2005). It is called the "gold standard" because it is the end time expected to win the gold medal at the next Olympic Games.¹ Gold standards allow to compare boats of different categories. If funding is not available for all categories, this is important for selecting boats for international competitions like the Olympic Games. Therefore, the subsequently used variable relative end time (REL_ENDTIME) is defined as follows (A_ENDTIME denoting absolute end time):

$$\text{REL_ENDTIME} = \text{A_ENDTIME} / \text{gold standard in the respective event}$$

This standardisation allows a direct comparison of the dependent variable across events. Table 1 shows descriptive statistics for all variables introduced above.

¹ On inquiry the German Rowing Association stated that the teams for the Beijing 2008 Games were selected using the times published in Kleshnev (2001). Only for two events the "gold standard" was adjusted to account for speed developments since Sydney 2000.

Table 1: Descriptive statistics of the rowing teams at the Olympic Games 2000

| Variable | Definitions | MV | SD | Min | Max |
|-----------|--|---------|---------|---------|--------|
| A_HEAT | Absolute end-time in heats (sec.) | 407.37 | 38.67 | 332.85 | 508.64 |
| A_SEMI | Absolute end-time in semi-finals (sec.) | 398.22 | 34.72 | 344.08 | 472.91 |
| A_FINAL | Absolute end-time in finals (sec.) | 399.31 | 36.45 | 333.08 | 492.84 |
| REL_HEAT | Relative end-time in heats | 1.0829 | 0.0274 | 1.0341 | 1.1835 |
| REL_SEMI | Relative end-time in semi-finals | 1.06536 | 0.0237 | 1.0258 | 1.1986 |
| REL_FINAL | Relative end-time in finals | 1.06253 | 0.02432 | 1.02897 | 1.2382 |
| HET | Tournament stage (HEAT = 1; SEMI = 2; FINAL =3) | 2 | 0.81 | 1 | 3 |
| SEX | Dummy (0 = female; 1 = male) | 0.60 | 0.49 | 0 | 1 |
| EXP | Average racing experience of the team (in years) | 6.4 | 2.6 | 0 | 14 |
| AGE | Average age of the team (in years) | 27.2 | 3.1 | 20 | 37 |
| WM99 | Rank achieved by the team at the 1999 rowing world championships | 8.5 | 6.3 | 1 | 20 |
| BOOT | Boat type (1-6) | 2.9 | 1.6 | 1 | 6 |
| LW | Lightweight category (dummy variable: 0= no; 1 = yes) | 0.28 | 0.45 | 0 | 1 |

Comparing the mean values already indicates that end times improve with progression in the tournament.² Despite the fact that the fastest mean times were rowed in the semi-finals, the difference between heats and finals (8.06 sec.) is statistically significant at the 5%-level ($t = 0.047^{**}$).

3. Estimation methods and empirical results

As discussed before, the data can be set in form of a balanced panel based on the tournament level. Therefore, the estimation is carried out using a random-effects linear regression model that controls for unobservable team-specific effects. The decision whether to use a random- or a conventional OLS model was taken based on the Breusch-Pagan Lagrange Multiplier Test (Breusch/Pagan 1979). Given that almost all of the independent variables are time-invariant (no variance within), it is not possible to apply an alternative fixed-effects specification, since all of them have been automatically dropped during the estimation process (Frick et al. 2009). This decision is confirmed by a significant ($\chi^2 = 64,08^{***}$) Lagrange-Multiplier Test (OLS vs. Random-Effects). Furthermore, a random-effects model also accounts for potential individual effects resulting from a variety of other non-observable and random variables (Matyas/Sevestre 1996: 94).

Hence, the estimation model takes the following form:

$$\text{REL_ENDTIME}_{ij} = \alpha_0 + \alpha_1 \text{EXP} + \alpha_2 \text{EXP_2} + \alpha_3 \text{AGE} + \alpha_4 \text{AGE_2} + \alpha_5 \text{WM99} + \alpha_6 \text{LW} + \alpha_7 \text{HET} + \alpha_8 \text{HET_2} + \alpha_9 \Sigma \text{BOOT}^3 + \varepsilon$$

² We only consider teams that have qualified for the final A or final B. Therefore, the quicker times in the finals cannot be the result of slower teams being excluded from the tournament.

³ ΣBOOT is a vector of six different boat types. Boat category 1 is 1x = single sculls; category 2 is 2x = double sculls; category 3 is 4x = quadruple sculls; category 4 is 2- = pair; category 5 is 4- = coxless four; and category 6 is 8+ = eight with coxswain.

Table 2 shows the estimation results of four different specifications; the calculations vary by the used estimation model and absolute vs. standardised dependent variable. Model 2 presenting the random-effects (RE) estimation for relative end-time is the preferred version; these results are the basis of the subsequent discussion. The other three model specifications give evidence for the robustness of our findings.

Table 2: Determinants of rowing intensity

| Variable | Model 1 (lnENDTIME) | | Model 2 (REL_ENDTIME) | |
|---------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| | Pool (OLS) | Random | Pool (OLS) | Random |
| EXP | -0.00424 (-2.44)*** | 0.00406 (-1.77)* | -0,00464 (-2,39)*** | -0,00437 (-1,67)* |
| EXP_2 | 0.00020 (1.76)* | -0.00019 (1.26)+ | 0,00021 (1,64)* | 0,00019 (1,13)+ |
| AGE | 0.00320 (0.96)+ | 0.00412 (0.91) | 0,00333 (0,89)+ | 0,00420 (0,83)+ |
| AGE_2 | -0.00004 (-0.90)+ | -0.00007 (-0.91)+ | -0,00006 (-0,85)+ | -0,00007 (-0,82)+ |
| WM99 | 0.00109 (5.57)*** | 0.00117 (4.78)*** | 0,00112 (5,04)*** | 0,00121 (4,37)*** |
| SEX | -0.09682 (-44.44)*** | -0.09759 (-33.57)*** | -0,00161 (-0,66)+ | -0,00249 (-0,77)+ |
| LW | 0.00501 (2.00)** | 0.00507 (1.47)+ | 0,01478 (5,22)*** | 0,01497 (3,80)*** |
| HET | -0.02996 (-3.56)*** | -0.02449 (-3.82)*** | -0,03256 (-3,53)*** | -0,02625 (-3,79)*** |
| HET_2 | 0.00501 (2.48)** | 0.00380 (2.40)** | 0,00562 (2,48)** | 0,00406 (2,38)** |
| BOAT1 | 0.22114 (57.78)*** | 0.22169 (42.92)*** | 0,02134 (5,16)*** | 0,02189 (3,87)*** |
| BOAT2 | 0.14505 (41.69)*** | 0.14498 (30.74)*** | 0,02412 (6,43)*** | 0,02422 (4,68)*** |
| BOAT3 | 0.047674 (13.67)*** | 0.04777 (10.44)*** | 0,00927 (2,47)*** | 0,00960 (1,92)* |
| BOAT4 | 0.16967 (45.42)*** | 0.16875 (32.69)*** | 0,01602 (3,95)*** | 0,01466 (2,59)** |
| BOAT5 | 0.08214 (22.86)*** | 0.08174 (17.36)*** | 0,01713 (4,42)*** | 0,01678 (3,27)*** |
| BOAT6 | REFERENCE | | REFERENCE | |
| CONST. | 5.91 (126.61)*** | 5.90 (93.33)*** | 1,05 (20,27) | 1,04 (14,78) |
| R ² | 0.957 | 0.957 | 0,371 | 0,370 |
| Adj. R ² | 0.955 | / | 0,344 | / |
| F-Value | 853.84*** | / | 18,65*** | / |
| Wald χ^2 | / | 6868.65*** | / | 295,59*** |
| Breusch-Pagan LM | 59.20*** | | 64.08*** | |
| N of Cases | 442 | | 442 | |

* p<0.10; **p<0.05; ***<p<0.01; + n.s. (White 1980 robust t und z-values in brackets).

All independent variables have the expected effect on end times; all coefficients possess the expected sign and lie within the statistical confidence intervals. The explanation of variance in absolute end-time is higher than 95%; this is in accordance with findings from other endurance sports analysing end times (Frick/Klaeren 1997). On the other hand, this result should be interpreted with caution. 58% of end-time variance is explained solely by the number of rowers in the boat (variables BOAT). Controlling for the categorical variables sex and lightweight, the eight is faster than all other boat types; singles and pairs are the slowest boats. In other words: more rowers make the boat faster. This dominant effect may bias or cover up the hypothesised effect of a heterogeneous competitor field. Therefore, standardising for boat types by gold standards, as it is common in rowing, is a useful measure for our investigation. This is confirmed by the results of model 2; the adjusted R^2 decreases by more than 50% (Adj. $R^2 = 0,34$), but all coefficients keep the expected algebraic sign.

Our analysis focuses on the hypothesised effect of heterogeneity (HET). Model 2 shows a significant negative coefficient; this indicates faster times in later, more homogeneous stages of the regatta. Statistically, the positive sign of the squared heterogeneity term (HET_2) counteracts this positive effect. However, this can be explained through exhaustion of all athletes at the end of the tournament (Prinz 2008). After all, our results confirm the contamination hypothesis prominent in the tournament literature: in heterogeneous competition the available price mechanisms do not have the same incentive effect on participants as in homogeneous competition. On average, contestants hold back effort in tournaments with heterogeneous line-ups.

The observed effects of our control variables are intuitively explained. In Model 1 (random-effects specification) the variable SEX indicates on average 40 sec faster races for the men's events compared to respective women's events. Similarly, because of their lighter physique, lightweight rowers (LW) are significantly slower than their heavyweight counterparts. Initially, more experienced crews (EXP) row faster times. With increasing age this effect diminishes. Hence, at later career stages additional experience does not outweigh deterioration in fitness. The positive coefficient of the squared experience term (EXP_2) yields a convex experience-power-profile with its minimum (i.e. maximum strengths) at the experience of 11 years, after that, athletes slow down (again). The rank at the 1999 world championship (WM99) has the expected positive coefficient. Each better rank yields – ceteris paribus – a 0.1% faster performance at the Olympic tournament. Hence, the rank achieved at the 1999 world championship is a good indicator for the skills and fitness of rowing teams at the Sydney games in 2000.

In an attempt of offering further evidence of our heterogeneity variable we re-estimate the random effects version of Model 2 (table 2) and substitute our linear experience parameter (EXP) by a simple binary variable (EXP_Dummy; EXP_D) taking on the values 0 for inexperienced and 1 for experienced athletes (Random-Effects-Alternative Model). This is advisable since too many inexperienced rowers (0-values) might bias our findings. Moreover, we use a “de-pooling” strategy by presenting the influence of our heterogeneity variable on the rowers' finishing times by splitting the sample into the six boat type categories.

Taken the results together, we contend that the hypothesised effect is undoubtedly confirmed in our data. Although the findings regarding the six different boat types of table 3 should be interpreted cautiously due to some dropped out variables (multicollinearity; less variance as well as number of cases) we find that – on average – heterogeneous competitions are rowed with less intensity.

Table 3: Robustness Check of rowing intensity; alternative models[#]

| Variable | RE-Alternative | BOAT1 | BOAT2 | BOAT3 | BOAT4 | BOAT5 | BOAT6 |
|---------------------|------------------------------------|------------------------------------|------------------------------------|----------------------------------|------------------------------------|------------------------------------|----------------------------------|
| EXP_D | -0.021 (-1.68)* | -0.118 (-0.65)+ | -0.026 (-2.43)** | dropped | dropped | dropped | dropped |
| AGE | 0.000 (0.10)+ | 0.001 (0.21)+ | 0.008 (1.14)+ | -0.049 (-1.91)* | -0.012 (-0.92)+ | 0.002 (0.24)+ | 0.063 (1.34)+ |
| AGE_2 | -0.000 (-0.18)+ | -0.000 (-0.24)+ | -0.000 (1.24)+ | 0.000 (1.88)* | 0.000 (0.76)+ | -0.000 (-0.31)+ | -0.001 (-1.33)+ |
| WM99 | 0.001 (4.61)*** | 0.002 (3.89)*** | 0.002 (6.70)*** | 0.002 (4.00)*** | -0.000 (-0.94)+ | 0.000 (1.63)+ | -0.000 (-0.24)+ |
| SEX | -0.004 (-1.37)+ | 0.009 (1.71)* | -0.014 (-3.59)*** | -0.019 (-3.79)*** | 0.010 (1.91)* | dropped | -0.001 (-0.22)+ |
| LW | 0.015 (4.00)*** | dropped | 0.025 (6.33)*** | dropped | dropped | -0.007 (-1.82)* | dropped |
| HET | -0.026 (-3.75)*** | -0.011 (-3.44)*** | -0.010 (-5.21)*** | -0.005 (-1.97)* | -0.015 (-5.58)*** | -0.012 (-5.76)*** | -0.000 (-0.19)+ |
| HET_2 | 0.004 (2.33)** | 0.003 (1.98)* | 0.003 (-1.99)* | 0.000 (1.13)+ | 0.005 (2.04)** | 0.004 (-1.99)+ | 0.000 (0.32)+ |
| BOAT1 | 0.019 (3.48)*** | / | / | / | / | / | / |
| BOAT2 | 0.024 (4.53)*** | / | / | / | / | / | / |
| BOAT3 | 0.009 (1.69)* | / | / | / | / | / | / |
| BOAT4 | 0.013 (2.34)** | / | / | / | / | / | / |
| BOAT5 | 0.016 (3.21)*** | / | / | / | / | / | / |
| BOAT6 | REFERENCE | / | / | / | / | / | / |
| CONST. | 1.102 (15.14)*** | 1.071 (9.60)*** | 1.021 (10.22)*** | 1.739 (5.07)*** | 1.300 (6.93)*** | 1.066 (7.92)*** | 0.181 (0.78)+ |
| R ² | 0.356 | 0.326 | 0.468 | 0.444 | 0.447 | 0.394 | 0.007 |
| Adj. R ² | / | 0.269 | 0.442 | 0.386 | 0.392 | 0.348 | -0.111 |
| Wald χ^2 | 292.18*** | / | / | / | / | / | / |
| N | 442 | 77 | 153 | 54 | 56 | 72 | 30 |

[#] "Boat estimations" are carried out running simples OLS-estimates. * p<0.10; **p<0.05; ***<p<0.01; + n.s.

Notice that the above models are not able to distinguish between favourites and underdogs. To think about measures reinstalling incentives, however, it is important to know whether both favourites and underdogs likewise hold back effort. In an effort to test whether this is indeed the case, we subsequently analyse race strategies and effort levels in a supplementary study focusing on the single sculling events.

4. Incentive effects of heterogeneous line-ups on favourites and underdogs

Rowing is an aerobic endurance sport that has been studied by scientists and coaches for a long time. Optimal racing strategies to complete the course in the fastest possible time have been developed and are well known among the athletes (Garland 2005; Teti/Nolte 2005). Hence, by comparing the split times for each of the 500m quarters, rowing experts can determine whether a team rowed the physiologically and psychologically optimal racing strategy or whether they held back effort during the course of the race.

Accelerating the boats from standstill to racing speed requires the highest effort level. However, because of the glycogen stored in the muscles, athletes can exceed the aerobic threshold at the beginning of a race without suffering an oxygen debt. Furthermore, rowers sit in their boats facing the stern; they see neither the finish line nor rivals ahead of them. Vice versa, the leader can observe his competitors without having to turn around; even for experienced rowers this would slow down the boat. Hence, despite the high effort required to accelerate the boat, rowers have good reason to start the race with the fastest split time. In the second quarter of the race, athletes must slow down to cruising speed in order to guarantee sufficient oxygen supply; otherwise lactic acid production would set in, the rowers would “die a slow death” on the course. Crews are advised to continue the same rhythms in the third quarter of the race. Neglecting the specifics of the human body’s energy supply system, even splits are the fastest race strategy from a pure hydrodynamic point of view. For the final sprint as part of the last quarter of the race, crews make use of anaerobic lactic energy supply and row higher speeds again. Hence, the ranking of splits in the “optimal racing strategy” is: first quarter, fourth quarter, second quarter, third quarter. Whenever the split for the fourth quarter is the slowest, the athletes have either deliberately slowed down or they misjudged their capacity. The latter is very unlikely to happen for experienced crews competing at Olympic Games. However, if it happens (as could be observed for New Zealand contender Mahe Drysdale in the Beijing 2008 men’s single scull final) it is accompanied by extraordinary fast splits in the second or third quarter. However, none such case was observed in the 2000 Sydney competition. Therefore, the subsequent analysis is based on the assumption that rowers deliberately hold back effort if the last quarter of the race is rowed in the slowest split time. At Olympic level, not rowing a final sprint is taken as a clear indicator of economising physical strengths.

Progression in the tournament is determined by the rank achieved in heats and repechages. Hence, if the ranks are taken by large margins at the 1.500m mark, there is no incentive for a favourite to increase his effort in the final quarter of the race. On the other hand, crews trailing behind are advised to show full effort in order to take advantage of any potential mishap in the boats of the leading crews. Athletes will economise on their strengths only if the price (progression in the tournament) is safe. This picture changes in the finals. First, there is no incentive to conserve energy for any further rounds of competition. Second, the strongest crews in the finals B, C, and D want to show by their end-time that they would have been able to compete in the

respective final one step further up. In the final A, even the gold medal favourite will only refrain from a final sprint if his position is absolutely unchallenged. The above considerations yield hypotheses 2 and 3, respectively:

Hypothesis 2: Favourites considerably hold back effort more often than underdogs.

Hypothesis 3: Athletes do significantly more often hold back effort in the preliminary stages of the tournament than in the final round.

The above discussed effects are much more difficult to observe in crew boats than in the single scull events. First, speed differences are smaller in crew boats; this makes it more difficult to interpret race strategies from the split times. Second, the effort shown by the individual athlete is primarily determined by the race strategy given by the coxswain or the crew member chosen to call for changes in boat speed. In general, tactically rowed races with deliberate slow-downs are less often observed in crew boats (Teti/Nolte 2005). For a first analysis, we therefore focus on the single scull events at the Sydney 2000 Olympics. In order to account for underdog specific effects, this time we include all entries in the analysis. Favourites and underdogs were coded by their final rank in the tournament, using a median split to divide the field. Since not all contenders finished all their races, the total sample consists of N=142 cases. To include deliberate holding back of effort in the analysis, we introduce the categorical variable shirking. As indicated before, by shirking we mean that a boat has slowed down in the final quarter of the race.

As predicted, table 4 shows correlations of the variable shirking with both variables favourite and final. Furthermore and not surprisingly, the variable favourite significantly correlates with experience.

Table 4: Descriptive statistics and correlations

| | mean | SD | 1. | 2. | 3. | 4. | 5. |
|----------------------------------|-------|------|---------|---------|----------|-------|------|
| 1. Age | 27.05 | 4.64 | 1 | | | | |
| 2. Experience | 5.56 | 3.58 | .548*** | 1 | | | |
| 3. Shirking (1 = yes) N = 141 | .53 | .501 | .185* | .201* | 1 | | |
| 4. Final (1 = yes) | .28 | .451 | -.024 | -.010 | -.469*** | 1 | |
| 5. Favourite (1 = yes) | .53 | .501 | .105 | .377*** | .231** | -.035 | 1 |
| 6. Sex (1 = male) | .56 | .499 | -.018 | -.028 | .158 | .024 | .036 |

N = 142, if not stated otherwise; * p<.05; ** p<.01; *** p<.001 (two-tailed)

Table 5 presents a cross-tabulation of the dependent variable shirking with both hypothesised variables (favourite and final). In 75 of 141 cases rowers showed deliberate holding back of effort; but only 6 of these cases were final round races. Only 27 of the races by underdogs were identified as deliberate hold-up. The table also shows that favourites had to row more races than underdogs; in order to qualify for the finals A and B contestants have to row semi-finals which are not needed to compete in the finals C and D. In addition, χ^2 -tests show that the differences in shirking for both favourites and underdogs as well as preliminary and final stages of the tournament are

statistically significant. This aligns with the contamination hypothesis (hypothesis 1) derived in section 3.

Table 5: Cross-tables and χ^2 -tests

| | Optimal strategy | Shirking | Total |
|-----------|------------------|----------|-----------------------|
| Favourite | 27 | 48 | 75 |
| Underdog | 39 | 27 | 66 |
| | 66 | 75 | $\chi^2 = 7.518^{**}$ |

| | Optimal Strategy | Shirking | Total |
|-----------|------------------|----------|-------------------------|
| Final | 33 | 6 | 39 |
| Non-final | 33 | 69 | 102 |
| | 66 | 75 | $\chi^2 = 30.950^{***}$ |

N = 141; * p<.05; ** p<.01; *** p<.001 (two-tailed)

Additionally, table 6 shows the results of a random-effects logistic regression for the dependent variable shirking. Final and favourite are taken as independent variables; we controlled for sex, age and experience of the athletes.⁴ The results clearly imply that neither hypothesis 2 nor hypothesis 3 can be refused. Surprisingly, the variable SEX also has a statistically significant effect at the 5%-level; male rowers are more likely to hold back effort than female athletes, a result that is opposite to the findings presented by Frick/Klaeren (1997). However, this effect may be due to the skewness of the end-time distribution among the competitors in the women's single sculling event. The field was dominated by the three medalists, namely Ekatarina Karsten-Khodotovitch (BLR), Rumyana Neykova (RUM), and Katrin Rutschow (GER). These three outstanding oarswomen passed the finish line within 9/10 of a second, but more than 8 seconds ahead of the rest of the field, whereas ranks 4 to rank 10 all achieved end-times within a span of 4 seconds. Hence, it is very likely that apart from the three medallists none of the other athletes coded as favourites by the median split, ever was in the comfortable position to deliberately slow down.

Table 6: Random-effects logistic regression model on shirking in rowing

| Dependent variable: Shirking (1 = yes; N = 141) | | | |
|---|----------------------|------|---------|
| Variable | Coefficient (random) | SE | z-value |
| Final (1 = yes) | -3.23 ^{***} | .700 | -4.62 |
| Age | .794 | .068 | 1.15 |
| Experience | .070 | .096 | .73 |
| Favourite (1 = yes) | 1.15 [*] | .592 | 1.94 |
| Sex (1 = male) | 1.07 [*] | .547 | 1.96 |
| Const. | -2.75 | 1.73 | -1.59 |

McFadden R² = .28

+ p < .1; * p < .05; ** p < .01; *** p < .001 (two-tailed)

⁴ Correlations between age, experience, and favourite indicate potential multi-collinearity. However, variance-inflation factors all do not exceed 2; the highest value being $VIF_{Age} = 1,76$. Hence, there is no multicollinearity between our independent variables.

5. Discussion

In our study of the Olympic Rowing Regatta in Sydney 2000 we found empirical evidence for the contamination hypothesis. On average, races in heterogeneous fields are rowed slower than races in close competition. In an additional study of the single sculling events, we deliver the first field evidence that favourites and underdogs react differently to heterogeneous fields. Whereas favourites take advantage of their strengths and hold back effort in preliminary stages of the tournament, underdogs significantly less often deliberately slow down their boats.

The results also provide evidence for the importance of the prize structure of rank-order tournaments. Whereas in preliminary stages athletes economise on their strengths and foremost secure progression, rowers show their best possible performance only in the finals. This has important implications for the use of rank-order tournaments in internal labour markets. Everyday job performance can not be modelled as a – in Olympics it maybe once in a lifetime – once only chance. Tournaments must remain a special event in order to unfold incentive effects. Therefore, tournaments should only be used to a limited degree. In everyday work life principals must consider supplementing tournaments by other incentive schemes that have less strict effect requirements than rank-order tournaments.

A general limitation to our study is the operationalisation of variables, in our case effort levels and heterogeneity. In the first part (section 3), taking the end time as an indicator for effort levels is an approach well known in the analysis of endurance sports. As expected, the results align with evidence from studies on other sporting events. However, the results could be distorted because of the variable tournament stage as a measure for heterogeneity. Despite a potentially heterogeneous line-up, each competitor starting in one of the six lanes in a rowing course may have one rival of similar strengths. Taken to an extreme, a heat may consist of three close matches set apart by large margins between rival pairs. Hence, the likelihood of winning a rank by only marginally increasing effort depends on the existence of one close competitor and not on the heterogeneity of the field of six; this condition may be given as well in a heat as in a final. However, our results are stable across all four model specifications. This implies that the variable tournament stage can be interpreted as an indicator for heterogeneity.

In the second part focusing on the single sculling events (section 4) we use a different measure for effort, namely the fact whether a rower takes a final sprint for the line or not. Hence, of the two options for the favourite to take advantage of his superior strengths discussed in the introduction, our coding only covers one, namely slowing down once the progression in the tournament is secured. The other option, adjusting effort levels to the speed of slower competitors from the very start, is not captured. However, even loosing out on additional cases of hold-up, our results are statistically significant.

Aside from the above limitations, our results imply that firms organising heterogeneous tournaments have to find ways to reinstall incentives of the favourites. One alternative mentioned in the literature would be to handicap favourites. This, however, may be problematic due to labour law regulations. A better way to keep favourites' in-

centives high is probably to enrich the tournament with absolute performance standards. If the size of the winner prize depends on the winner's absolute performance (i.e. on whether the winner's performance is above some standard), favourites have an incentive to put forth effort even if they are far ahead of their competitors since slacking off may come at the risk of not meeting the performance standard.

Finally, the evidence that rowers do not hold back effort in the finals implies that awarding absolute achievements (i.e. end-time) has a more profound incentive effect than awarding rank-order. This implication is fundamental for firm internal incentive schemes like e.g. goal attainment of sales forces.

Summarising the above discussion, the achieved results show that analysing the prize structure of rank-order tournaments is a promising field for further empirical research. Furthermore, the sport of rowing proved to provide suitable data to test theoretically derived hypotheses. Especially for research questions regarding different prize structures, rowing with national associations who differ in their regatta regulations provides ample opportunity for future empirical work.

References

- Breusch, T./Pagan, A. (1979): A simple Test for Heteroscedasticity and Random Coefficient Variation. In: *Econometrica*, 47: 1287-1294.
- Baker, G./Gibbs, M./Holmström, B. (1994): The Wage Policy of a Firm. In: *Quarterly Journal of Economics*, 109: 921-955.
- Clark, D./Riis, C. (2001): Rank-order Tournaments and Selection. In: *Journal of Economics*, 73: 167-191.
- Ehrenberg, R./Bognanno, M. (1990): Do Tournaments Have Incentive Effects. In: *Journal of Political Economy*, 98: 1307-1324.
- Fair, R. (1994): How Fast do old Men Slow Down? In: *Review of Economics and Statistics*, 76: 103-118.
- Frick, B./Klaeren, R. (1997): Die Anreizwirkungen leistungsabhängiger Entgelte: Theoretische Überlegungen und empirische Befunde aus dem Bereich des professionellen Sports. In: *Zeitschrift für Betriebswirtschaft*, 67: 1117-1138.
- Frick, B./Gürtler, O./Prinz, J. (2008): Anreize in Turnieren mit heterogenen Teilnehmern – Eine empirische Untersuchung mit Daten aus der Fußball- Bundesliga. In: *Zeitschrift für betriebswirtschaftliche Forschung*, 60: 385-405.
- Frick, B./Gürtler, O./Prinz, J., Wiendl, A. (2009): Einkommens- oder Reputationsmaximierung? Eine empirische Untersuchung der Vergütung und Leistung von Bundesliga-Schiedsrichtern. In: *Die Betriebswirtschaft*, 69: 1-15.
- Garland, S. (2005): An analysis of the pacing strategy adopted by elite competitors in 2.000m rowing. In: *British Journal of Sports Medicine*, 39: 39-42.
- Hübl, L./Swieter, D. (2002): Der Spielermarkt der Fußball-Bundesliga. In: *Zeitschrift für Betriebswirtschaft, Ergänzungsheft Sportökonomie*, 72: 105-125.
- Kahn, L. (1993): Free Agency, Long Term Contracts and Compensation in Major League Baseball: Estimates from Panel Data. In: *Review of Economics and Statistics*, 75: 157-164.
- Kleshnev, V. (2001): Racing strategy in rowing during Sydney Olympic Games. In: *Australian Rowing*, 24: 20-23.
- Kräkel, M. (2002): U-Type versus J-Type Tournaments. In: *Journal of Institutional and Theoretical Economics*, 158: 614-637.
- Lazear, E./Rosen, S. (1981): Rank-Order Tournaments as Optimum Labor Contracts. In: *Journal of Political Economy*, 89: 841-864.
- Lazear, E. (1989): Pay Equality and Industrial Politics. In: *Journal of Political Economy*, 97: 561-580.
- Lynch, J. (2005): The effort effects of prizes in the second half of tournaments. In: *Journal of Economic Behavior & Organization*, 57: 115-129.

- Malcomson, J. (1984): Work Incentives, Hierarchy, and Internal Labor Markets. In: *Journal of Political Economy*, 92: 486-507.
- Matyas, L./Sevestre, P. (1996): *The Econometrics of Panel Data*. Dordrecht/Boston/London: Springer.
- Maxcy, J. (1996): Do long-term Contracts influence Performance in Major League Baseball. In: *Advances in the Economics of Sport*, 2: 157-176
- Meyer, M. (1991): Learning from Coarse Information: Biased Contests and Career Profiles. In: *Review of Economic Studies*, 58: 15-41.
- Prinz, J. (2008): Gibt es einen Trade-Off zwischen Skilauf- und Schießleistung? Erscheint in: M. Büch/H. Dietl/E. Franck/H. Kempf (Eds.): *Fußball: Ökonomie einer Leidenschaft*. Beiträge der 11. Fachtagung des Arbeitskreises Sportökonomie. Schorndorf: Hofman, 2008.
- Rajan, M./Reichelstein, S. (2006): Subjective Performance Indicators and Discretionary Bonus Pools. In: *Journal of Accounting Research*, 44: 585-618.
- Rosen, S. (1986): Prizes and Incentives in Elimination Tournaments. In: *American Economic Review*, 76: 701-715.
- Schotter, A.; Keith W. (1992): Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results. In: *Quarterly Journal of Economics*, 107: 511-539.
- Teti, M./Nolte, V. (2005): Setting Race Plans and Tactics. In: V. Nolte (Ed.): *Rowing Faster*. Champaign/Il.: Human Kinetics Publisher: 237-248.
- White, H. (1980): A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. In: *Econometrica*, 48: 817-838.